

**STRATEGIEN GEGEN AUSREISSER  
IN ZEITREIHENMODELLEN**

Robert Kunst

Forschungsbericht/  
Research Memorandum No. 219

Juni 1985

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

### Zusammenfassung

Die vorliegende Arbeit versteht sich nicht nur als Korrektur und Erweiterung zur Präsentation eines Zeitreihenmodells der österreichischen Wirtschaft in: Kunst (1985), sondern möchte allgemeine Strategien anbieten, um den Einfluß von Ausreißern in univariaten und eventuell auch multivariaten Zeitreihenmodellen unter Kontrolle zu bringen. Der Schwerpunkt liegt hier auf automatischen Verfahren, für welche eine Kenntnis der Art und der Zeitpunkte der Ausreißer überflüssig ist.

Im zweiten, empirischen Teil der Arbeit werden zwei der geschilderten Methoden auf ökonomische Daten angewendet, und zwar rein univariat. Es ergibt sich hinreichende Evidenz für einen nicht zu vernachlässigenden Einfluß beider gebräuchlichen Spezifikationen von Ausreißern (additive und innovations outlier) auf die Modellschätzung und auf die Prognose.

### Abstract

This paper is to be understood not only as correction and extension of the older paper by the same author which presented a time series model of the Austrian economy, it also wants to offer strategies to control the influence of outliers in univariate as well as multivariate time series models generally. The stress is on automatic procedures which do not require knowledge a priori concerning origin or location in time of the outliers.

In the second part of the paper, two procedures are applied to economic data in a purely univariate framework. Sufficient evidence is found for a non-negligible influence of both innovations and additive outliers on model estimation and forecasting.



# INHALTSVERZEICHNIS

Seite

Einleitung	1
1. Elemente	2
1.1 Ausreißermodelle	2
1.2 Robustes Filtern	5
1.3 Identifikation der Fehlerverteilung	9
1.4 Adaptive stochastische Prognose	13
2. Durchführung	16
2.1 Daten	16
2.2 Software	16
2.3 Erste empirische Evidenz: robuste Spektren	17
2.4 Versuch einer Interpretation der Spektren	20
2.5 Verteilungsschätzung anhand der Rohdaten	22
Literatur	24
Anhang	



## Einleitung

Wie überall in der Statistik, wo mit quadratischen Kriterien gearbeitet wird, so gilt auch bei der Schätzung von Zeitreihenmodellen, daß einzelne Beobachtungen, deren Verhalten mit jenen der übrigen nicht übereinstimmt, die gesamte Schätzung aus den Angeln heben können. Treten nun solche "Ausreißer" auf, so hat der Modellbauer mehrere Möglichkeiten zu reagieren.

Die in der Ökonomie verbreitetste Reaktionsweise besteht im "Kaltstellen" der bösartigen Beobachtung. Dies erfordert eine Identifikation der Ausreißer ebenso wie eine Rechtfertigung dafür, gerade diese Daten aus der Stichprobe herauszunehmen. Die Identifikation der Ausreißer ist in multivariaten Modellen überraschend komplex. Hat man nicht zufällig die wahren Ursachen des Ausreißers bei der Hand, so ist meist schwer zu begründen, warum ein bisher schon gelegentlich ungültiges Modell in Hinkunft immer gelten soll. Jegliche Asymptotic kann dann eigentlich vergessen werden.

Eine andere Reaktion wäre das Aufgeben des Modells bzw. die deskriptive Anpassung an das Sample. Was für ein Gefühl hat aber der Laie, z.B. ein Autofahrer, dessen Auto neunmal 10 Liter und einmal 20 Liter Treibstoff verbraucht? Würde er wirklich 11 Liter für die nächste Tankfüllung als Verbrauch prognostizieren? Wie wäre seine Prognose, wenn 20 Liter die erste Beobachtung wäre und wie, wenn die letzte?

Die robuste Statistik arbeitet im wesentlichen mit einer Änderung des Gütekriteriums. Statt quadratischer Abweichungen werden etwa absolute Abweichungen oder andere Funktionen verwendet (Huber's  $\rho$  und  $\psi$ -Funktionen). Ähnlich wie sich quadratisches Kriterium und Normalverteilung gegenseitig implizieren, so liegen auch für die anderen Kriterien passende Verteilungen vor. Meist zeigt sich, daß ein solches Vorgehen zu ganz ähnlichen Resultaten führt, wie der Dummy-approach, jedoch theoretisch viel sauberer ist. Ein Problem ist dann die richtige Kriteriumsfunktion und natürlich, ob diese "richtige" eventuell tatsächlich die Verlustfunktion des Prognostikern wiedergibt (d.h. seine Kosten bei Fehlprognose).

## 1. Elemente

### 1.1 Ausreißermodelle

In der neueren Literatur über robuste Zeitreihenanalyse unterscheidet man zwischen zwei wesentlich verschiedenen Ausprägungen des Ausreißerverhaltens. Diese Einteilung geht auf Fox (1972) zurück, eine der ersten Arbeiten auf diesem Gebiet überhaupt.

- a) Die Beobachtung zum Zeitpunkt  $t$  weicht vom Grundmodell ab. In das Verhalten der Zeitreihe in  $t+1, t+2, \dots$  geht aber die dem Grundmodell entsprechende, nicht beobachtbare Realisierung ein. Man spricht vom AO-(additive outliers)-Modell. Es besteht ein offensichtlicher Bezug zur errors-in-variables-Problematik.
- b) Die Beobachtung zum Zeitpunkt  $t$  weicht vom Grundmodell ab. Das Verhalten der Zeitreihe zu späteren Zeitpunkten erklärt sich aus der gestörten Beobachtung, womit auch die nachfolgenden Beobachtungen gedämpfte Abweichungen zeigen. Man spricht vom IO(innovations outliers)-Modell, da es hier logisch erscheint, das Ausreißerverhalten in den Innovationsprozeß zu integrieren.

In Formeln läßt sich AO so anschreiben:

$$y_t^* = A(L)e_t$$

$$y_t = y_t^* + f_t$$

$y_t^*$  ist hier der systeminterne, nicht beobachtbare Wert von  $y$ ,  $f_t$  ist der "Nebel", durch den der Beobachter das  $y$  wahrnimmt. Im allgemeinen wird  $f_t$  iid in Form einer Mischverteilung  $P_f = \lambda I_0 + (1-\lambda)P$  angenommen, wobei  $I_0$  die in 0 konzentrierte Punktverteilung und  $P$  irgendeine, relativ zum Gesetz der  $e$  flache, in 0 zentrierte Verteilung repräsentiert. Diese Spezifikation widerspiegelt das ursprüngliche Konzept von Meß- und Übertragungsfehlern.  $\lambda$  sollte dementsprechend hoch, etwa bei .9 oder .95, liegen. Die beste Prognose für  $y_t$  ergibt sich unter AO aus

$$E(y_t/y_{t-1}^*, y_{t-2}^*, \dots) = E(y_t^*/y_{t-1}^*, y_{t-2}^*, \dots)$$



wogegen selbst bei korrekter Spezifikation der Transferfunktion  $A(L)$  bei der üblichen Prognoseform

$$E(y_t/y_{t-1}, y_{t-2}, \dots) = E(y_t^*/y_{t-1}, y_{t-2}, \dots)$$

eine nichttriviale Verzerrung auftritt. Die gebräuchlichen Methoden zielen nun alle darauf hin,  $y_t^*$  möglichst gut zu identifizieren. Zu diesem Zweck werden entweder Beobachtungen mit hohen Residuen

$$y_t - \hat{A}(L)\hat{e}_t = f_t + (A - \hat{A})(L)e_t + \hat{A}(L)(e_t - \hat{e}_t) = \hat{f}_t$$

geringer gewichtet oder um das  $\hat{f}_t$  korrigiert. Man sieht an diesem Ausdruck, daß sich der Schätzwert  $\hat{f}$  aus drei Einflüssen zusammensetzt, nämlich dem wahren Störprozeß, dem Schätzfehler in bezug auf die Transferfunktion und dem damit zusammenhängenden Schätzfehler in bezug auf die Innovationen. Abgesehen von eventuellen Fehlspezifikationen und small-sample-Abweichungen gilt demnach, daß bei großem  $f$  über den Bias bei der Berechnung von  $\hat{A}$  ein großer Unterschied zwischen  $f$  und  $\hat{f}$  auftritt, Korrekturen oder Umgewichtungen also wieder ziemlich unexakt sind. Der iterative Filter nach Thomson (1977) hofft darauf, daß die Abweichungen zwischen  $A$  und  $\hat{A}$  unter Zuhilfenahme der  $y_t - \hat{f}_t$  geringer sind als die zwischen  $A$  und  $\hat{A}$  unter Zuhilfenahme der  $y_t$ , womit die neugeschätzten  $\hat{f}_t$  noch näher an  $f_t$  liegen, kurzum, daß das Verfahren iterativ konvergiert. Schlechte Startwerte oder Fehlspezifikationen von  $A, P_f$  könnten hier gefährlich werden.

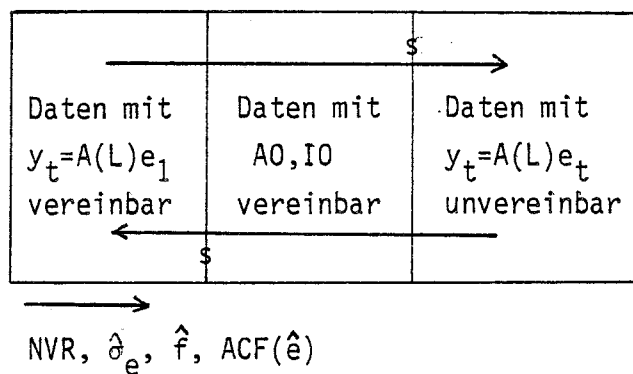
IO schreibt sich wie ein gewöhnliches Zeitreihenmodell

$$y_t = A(L)e_t$$

nur daß die Verteilung  $P_e$  von  $e_t$  "non-standard" ist, etwa  $P_e = \lambda N(0, \sigma^2) + (1 - \lambda)N(0, k\sigma^2)$  mit z.B.  $\lambda = .9$  und  $k=10$  oder  $P_e = S_\alpha$  mit  $\alpha < 2$  ( $S_\alpha$  steht im folgenden für stabile Verteilungsgesetze mit stabilem Index  $\alpha$ ). Die Güte der Prognose wird hier lediglich durch den Schätzfehler für  $A$  bestimmt, hängt also von der "distributional robustness" des Schätzverfahrens ab. In vielen Fällen ist diese überraschend gut (siehe Hannan & Kanter (1977) oder Kunst (1984) für weitere Referenzen). Die Varianz des Prognosefehlers ist in der Form von  $P_e$  begründet

und zumeist entsprechend groß. Abgesehen vom Falle nichtzentraler  $P_e$  gibt es keine Möglichkeit, den Prognosefehler zu reduzieren. Eher als in der Auffindung besserer Punktschätzer liegt also hier die Aufgabenstellung in der passenden Spezifizierung von Konfidenzbereichen.

Die Ausreißermodelle sind erweiterbar, etwa durch Zulassen von Autokorrelation in den Störungen. Prinzipiell wird durch die Ausreißermodelle eine Grauzone abgesteckt zwischen dem Bereich, in dem das Grundmodell ( $y_t = A(L)e_t$ ) gilt und demjenigen, in dem es nicht gilt.



Dieses schematische Diagramm soll die Situation veranschaulichen. Aufgrund gewisser goodness-of-fit-Werte lassen sich drei Bereiche unterscheiden. Im ersten paßt das vorgegebene Modell, im dritten wird es statistisch verworfen, im mittleren läßt es sich unter Kompromissen aufrechterhalten. Die Pfeile mit den S signalisieren das katastrophentheoretische Verhalten des Analytikers, der, von links kommend, das Modell in S endgültig verwirft, von rechts kommend aber erst in S genügend Evidenz zur Annahme des Modells vorfindet. Am Fuße des Diagramms erscheinen die hauptsächlichen goodness-of-fit-Statistiken: der Anteil der Residuen an der Gesamtvarianz, der absolute Schätzwert für die residuale Varianz, die Zeitreihe der geschätzten Prognosefehler und die Autokorrelation der Residuen. Ein großes Vertrauen in die Existenz eines Zeitreihenmodells kann den mittleren Bereich sehr weit nach rechts ausweiten.

Man überlegt sich leicht, daß im statischen Fall IO- und AO-Modell zusammenfallen bzw. sich nur in der Spezifikation der Fehlerverteilung unterscheiden. Das Festhalten der IO-Modellierer an leptokurtischen Verteilungen, der AO-Modellierer an Mischverteilungen hat historische wie technische Gründe.

## 1.2 Robustes Filtern

Bereits in 1.1. wurden Methoden berührt, AO-Modellen zu Leibe zu rücken. Prinzipiell lassen sich die Verfahren in zwei Klassen einteilen:

- A) Verfahren, die direkt aus den Realisierungen des beobachteten Prozesses  $y_t$  auf das Verhalten des zugrundeliegenden Prozesses  $y_t^*$  zurückschließen, d.h. im üblichen Sinne gegen das Auftreten der Störungen  $f_t$  "robust" sind.
- B) Verfahren, die zunächst versuchen, über eine Schätzung des Störprozesses  $f_t$  den Prozeß  $y_t^*$  möglichst gut zu rekonstruieren und dann aus diesem rekonstruierten Prozeß die Schätzung des Zeitreihenverhaltens ableiten.

Zu den am weitesten entwickelten Verfahren der Klasse A zählt das GM-Verfahren. Im Falle von Zeitreihenmodellen genügt es ja nicht, den Einfluß großer Residuen zu dämpfen, wie etwa bei der Lösung der M-Regression im statischen Fall (Modell  $y_t = x_t' \beta + e_t$ )

$$\sum_{t=1}^T \rho(y_t - x_t' \beta) \rightarrow \min \Leftrightarrow \sum_{t=1}^T x_t' \psi(y_t - x_t' \beta) = 0$$

Enthält nämlich hier der Vektor  $x_t$  die verzögerten Werte von  $y_t$  und sei  $y_s$  eine gestörte Beobachtung ( $f_s = 0$ ), so sind die folgenden Residuen (in  $s+1, s+2, \dots$ ) auch dann anomal, wenn  $f_{s+1}, f_{s+2}, \dots = 0$ . In der GM-Regression wird nun auch der Einfluß der Regressoren gedämpft:

$$\sum x_t' w(x_t) \psi(y_t - x_t' \beta, s(x_t)) = 0$$

$s(x)$  versteht sich als Skalierungsfunktion.  $w(x)$  ist eine Gewichtsfunktion, die großen  $x$  kleine Gewichte zuweisen soll. Als früheste Veröffentlichung zu diesen Schätzern scheint Kleiner, Martin & Thomson (1979) dazustehen. Das Konzept wurde an anderer Stelle, z.B. in Maronna, Bustos & Yohai (1979), verschiedentlich erweitert.

Im Zeitreihen-Kontext läßt sich jedes derartige Verfahren sofort auf AR- oder ARX-Modelle anwenden - genau das war die Idee bei Kleiner et al. - iterativ dementsprechend auch auf ARMA-Modelle. Der Verzicht auf die Identifikation der Ausreißer oder des zugrundeliegenden Prozesses führt aber bei

Prognoseanwendungen auf ein fundamentales Problem. Angenommen, die Transferfunktion  $A(L)$  wäre geschätzt. Verwendet man nun (in einleuchtender Notation)

$$E(y_t/y_{t-1}, y_{t-2}, \dots; \hat{A})$$

als Schätzer für das zu prognostizierende  $y_t$ , so ist die Prognose selbst dann schlecht, wenn  $A = \hat{A}$  genau vorliegt, solange unter  $y_{t-1}, y_{t-2}, \dots$  Ausreißer auftreten. Das GM-Verfahren ist also nur dann verwendbar, wenn zusätzlich einzelne Beobachtungen auf ihre Ausreißerwahrscheinlichkeit getestet werden. Eine Anwendung des Konzeptes für den statischen Fall (Cook (1979)) auf Zeitreihenmodelle behandelt Ledolter (1985). Die als Ausreißer erkannten Beobachtungen müßten dann ausgeglättet werden. Wie man sieht, kommt man auch hier um eine Untersuchung des  $f_t$ -Prozesses nicht herum.

Das am weitesten verbreitete Verfahren der Klasse B ist der robuste Filter nach Thomson (1977), später zum ACM-Filter verfeinert (Martin (1979)) und auf ARMA-Modelle erweitert (Martin, Samarov & Vandaele (1981)). Hier wird in einem ersten Schritt (STEP 1) die gesuchte Transferfunktion  $A(L)$  mittels eines herkömmlichen (nicht robusten) Verfahrens identifiziert. Im zweiten Schritt (STEP 2) werden, nach Berechnung einer Skalierung  $s$  (aus den Residuen, nicht direkt aus den Daten) die Daten  $y_t$  durch gewichtete Mittel aus Daten und Prognosewerten aus STEP 1 ersetzt. Im AR-Modell hat STEP 2 folgendes Aussehen (Grundmodell  $y_t = \sum A_j y_{t-j} + e_t$ ):

$$\hat{y}_t = \sum \hat{A}_j \hat{y}_{t-j} + s \psi((y_t - \sum \hat{A}_j \hat{y}_{t-j})/s), \quad t=p+1, \dots, T$$

Die  $\hat{y}_t$  sollten nun den  $y_t^*$  ähnlicher sein als die  $y_t$ . Man geht also mit den  $\hat{y}_t$  zu STEP 1 zurück und iteriert solange weiter, bis sich die  $\hat{A}_j$  nicht mehr ändern, womit man hofft,  $\hat{y}_t = y_t^*$  mehr oder minder erreicht zu haben.

Wo  $\psi(x) = x$  ist, bleiben die  $y_t$  erhalten. Unter der Annahme des AO-Modells sollte dies etwa in 90-95% der Fälle passieren. Bei der praktischen Anwendung des Verfahrens zeigt sich oft ein viel größerer Prozentsatz an korrigierten Daten, natürlich in Abhängigkeit von der gewählten  $\psi$ -Funktion. Verwendet man Huber's  $\psi$  mit Standard-Konstante, d.h.

$$\psi(x) = \min(|x|, 1) \cdot \text{sgn}(x)$$

so gilt  $\hat{y}_t = y_t$ , wenn

$$|y_t - \sum \hat{A}_j \hat{y}_{t-j}| \leq s$$

Eine sinnvolle Spezifikation für  $s$  wäre also ein Fraktil der unterlegten Verteilung der  $e_t$  (z.B. das 90%-Fraktil der  $N(0, \hat{\sigma}^2)$ ), wobei eventuell wieder gewisse Parameter dieser Verteilung aus STEP 1 zu bestimmen sind. Kleiner et al. (1979) behelfen sich mit

$$\text{med}(|\hat{y}_t - \sum \hat{A}_j \hat{y}_{t-j}|) / .6745$$

behalten sich aber die Möglichkeit vor, diesen Ausdruck mit einer "tuning constant" zu multiplizieren, um schnellere Konvergenz zu erzielen. Im später entwickelten ACM-Filter ist  $s$  überhaupt nicht mehr zeitinvariant. Bei Martin (1979) wird ein Hampel- $\psi$  gemeinsam mit speziellen Konstanten empfohlen.

Die auf dem robusten Filter aufbauende Prognose berechnet sich nun aus

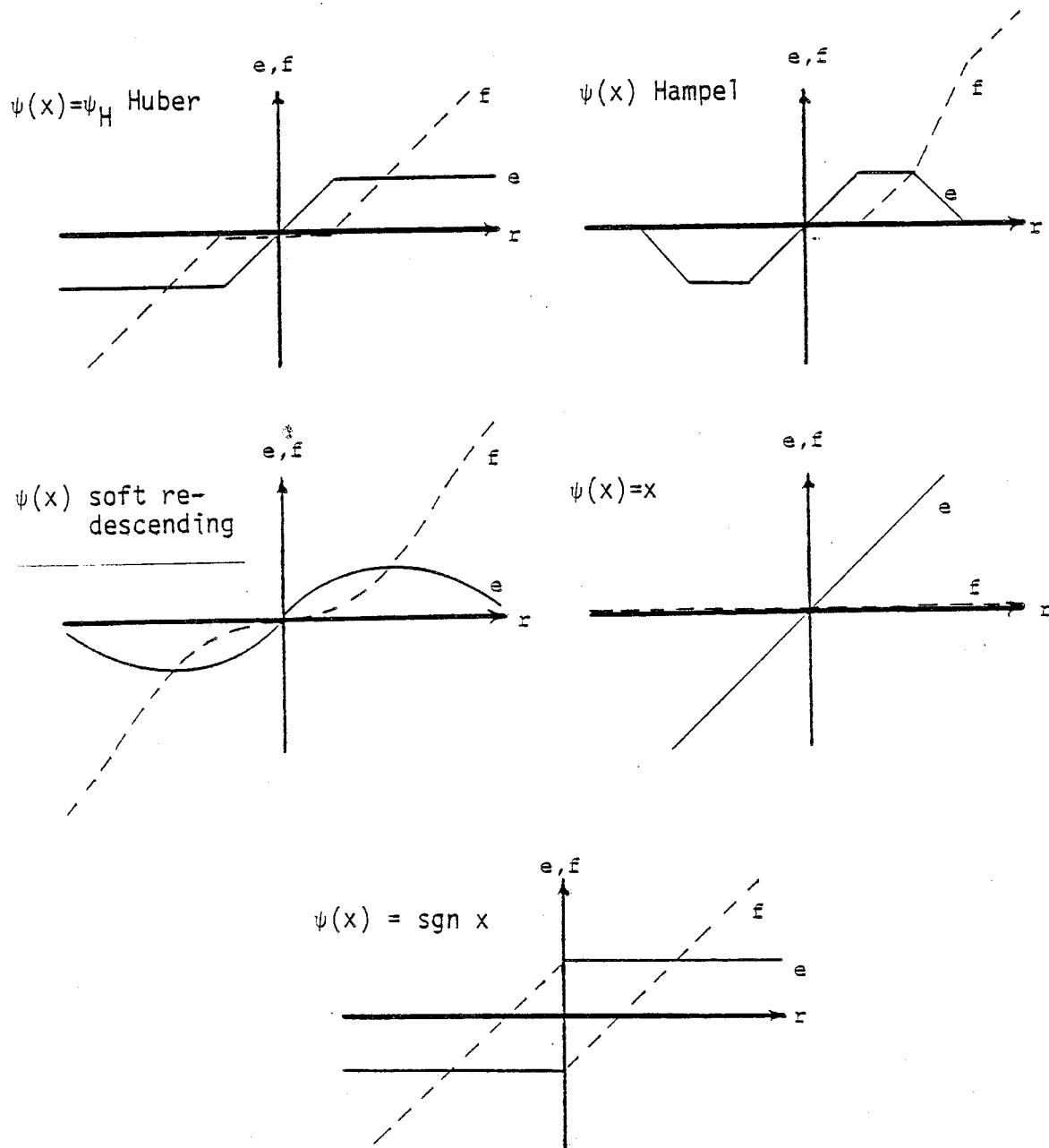
$$E(y_t / \hat{y}_{t-1}, \hat{y}_{t-2}, \dots; \hat{A})$$

direkt unter optimaler Ausnützung der zur Verfügung stehenden Information. Wichtig dürfte allerdings sein, darauf hinzuweisen, daß das Verfahren zur Korrektur einzelner Ausreißer ( $f_t$  iid) erschaffen wurde und nicht für "patchy outliers" geeignet ist. Es ist gut möglich, daß die dem Kalman-Filter näher stehenden robusten ACM-Versionen in dieser Richtung gewisse Vorteile aufweisen.

Ungeklärt ist auch die richtige Wahl der  $\psi$ -Funktion, die natürlich eng mit der den  $f_t$  unterlegten Verteilung zusammenhängt. Im Falle von Huber's  $\psi$  wird bei kleinen Residuen  $r_t$ , die ja einer Approximation von  $e_t + f_t$  entsprechen, so getan, als sei  $f_t = 0$ ; große Residuen werden wie  $e_t = \pm s$ ,  $f_t = r_t \mp s$  behandelt. Ist  $\psi$  hard-receding (z.B. Hampel's  $\psi$ ), so wird bei großen  $r_t$   $e_t = 0$  und  $f_t = r_t$  gesetzt. Ist aber  $e_t$  als normal- oder sonst stetig verteilt vorausgesetzt, so tritt  $e_t = 0$  mit Wahrscheinlichkeit 0 auf. In diesem Zusammenhang scheint Hampel's  $\psi$  bedenk-

lich. Es ist denkbar, daß ein "soft redescending" (d.h.  $\psi(x) \rightarrow 0$  für  $|x| \rightarrow \infty$ , aber  $\psi(x) \neq 0$  für  $x \neq 0$ ) für eine größere Klasse von AO-Verteilungsannahmen gute Eigenschaften besitzt. Martin (1981) verwendet im ACM-Filter eine Hampel-Funktion, weist aber darauf hin, daß diese im robusten AR-Filter dazu tendiert, einen überproportionalen Datenteil auszuglätten.

Diagramm: Aufteilung der Residuen auf e und f bei verschiedenen  $\psi$ -Funktionen



### 1.3. Identifikation der Fehlerverteilung

Prinzipiell ist damit zu rechnen, daß beide Formen von Ausreißern im vorliegenden Datensatz auftreten können, d.h. AO- und IO-Ausreißer. Da die statistischen Eigenschaften der Parameterschätzer durch AO beeinträchtigt werden, ist es nötig, AO-Korrekturen vor bzw. gleichzeitig mit der Parameterschätzung zu erledigen. Demgegenüber ist die Parameterschätzung durch IO größtenteils ungestört, in manchen Fällen tritt sogar eine starke Verbesserung der Konvergenzgeschwindigkeit ein (siehe Hannan & Kanter (1977), Yohai & Maronna (1977) sowie Kunst (1984) für den ARX-Fall und weitere Referenzen).

Sind die zweiten Momente der Fehlerverteilung unendlich, so verlieren herkömmliche Kriterien über Prognosegüte jedenfalls ihren Sinn. So ist etwa in

$$y_t = \alpha y_{t-1} + e_t, \quad E(e_t^2) = \infty, \quad E e_t = 0$$

die Prognose von  $y_t$  durch

$$E(y_t / \Omega) = \alpha y_{t-1} = \hat{y}_t \quad \text{mit } \Omega = [y_{t-1}, y_{t-2}, \dots]$$

wegen

$$E((y_t - \hat{y}_t)^2 / \Omega) = E e_t^2 = \infty$$

nicht mehr die beste Prognose im Sinn des quadratischen Kriteriums. Andererseits ist es unmöglich, einen Prognosewert zu finden, der

$$\hat{y}_t \in \Omega, \quad E(y_t - \hat{y}_t)^2 / \Omega < \infty$$

erfüllt wegen

$$\begin{aligned} E(y_t - \hat{y}_t)^2 / \Omega &= E(\alpha y_{t-1} + e_t - \hat{y}_t)^2 / \Omega = E e_t^2 / \Omega + E(y_{t-1} - \hat{y}_t)^2 / \Omega + \\ &+ E(e_t(\alpha y_{t-1} - \hat{y}_t))^2 / \Omega = \infty \end{aligned}$$

IO-Modelle solcher Art sind also ausgezeichnet durch

- a) sehr schnelle Konvergenz von  $\hat{\alpha}_T \rightarrow \alpha$ , allgemein  $\hat{A}(L) \rightarrow A(L)$
- b) sehr große Prognosefehler, die sich nicht verringern lassen

In diesem Sinne ist das einzige, auf das wir hoffen können, ein möglichst guter Einblick in das Verhalten des Prozesses  $e_t$ , im besten Fall dessen genaue Verteilungsfunktion. Sodann können wir durch stochastische Prognose auch dem Einfluß auf etwaige sonstige Variable (in multivariaten Modellen) ein Stück näher kommen.

Unterlegt man für  $(e_t)$  eine Normalverteilung  $N(0, \sigma^2)$ , so lassen sich stochastische Prognosen durch Zufallszahlen mit Verteilung  $N(0, \frac{\sigma^2}{n})$  durchführen, es ist also lediglich ein Parameter aus den Residuen zu schätzen. Unterlegt man eine größere Familie von Verteilungen, z.B. t-Verteilungen, so sind meist weitere Parameterschätzer nötig, i.a. eine Skalierung und ein Index für das asymptotische Verhalten der Dichte (bei den t-Verteilungen etwa die Freiheitsgrade). Mit zunehmender Stichprobengröße verbessern sich diese Parameterschätzer. Eine nichtparametrische Dichteschätzung kann nur bei sehr großen Stichproben sinnvoll sein.

Die Wahl der Verteilungsfamilie ist eher Geschmackssache. Sie sollte folgende Anforderungen erfüllen:

- a) die Normalverteilung (d.i. das Standardmodell) sollte enthalten sein.
- b) sie sollte von wenigen Parametern abhängen; diese sollten leicht schätzbar sein.
- c) Zufallszahlen mit den entsprechenden Verteilungen sollten ohne großen Aufwand generierbar sein.
- d) sie sollte nicht unplausibel sein.

Drei Familien können daher in erster Linie ins Auge gefaßt werden:

1. die Familie der e.p.-Verteilungen mit Dichten von der Form  $s(c, \alpha) \cdot \exp(-c|x|^\alpha)$   
( $\alpha=2$  ergibt die Normalverteilung,  $\alpha$  und  $c$  sind zu schätzen)
2. die Familie der t-Verteilungen  $T(n, c)$   
( $n \rightarrow \infty$  ergibt die Normalverteilung,  $n$  und  $c$  sind zu schätzen)
3. die Familie der stabilen Verteilungen mit charakteristischen Funktionen der Form

$$\exp\{-\gamma|t|^\alpha[1+i\beta \operatorname{sgn}(t)\omega(t, \alpha)]\} \quad \omega(t, \alpha) = \begin{cases} \tan(\pi\alpha/2) & \alpha \neq 1 \\ 2/\pi \ln|t| & \alpha = 1 \end{cases}$$

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad \gamma > 0$$

( $\alpha=2$  ergibt die Normalverteilung;  $\alpha, \beta, \gamma$  sind zu schätzen)



Die dritte Familie bietet den Vorteil, mit dem Argument zentraler Grenzwertsätze (Störung als Summe vieler Einflüsse) verträglich zu sein und Verteilungen mit sehr hohem Ausreißeranteil (z.B. Cauchyverteilung für  $\alpha = 1$ ) ebenso zu berücksichtigen wie die Normalverteilung. Ihr soll im weiteren das Hauptaugenmerk gelten. Außer den angegebenen Verteilungsfamilien sind auch Mischverteilungen wie im AO-Fall (etwa  $\lambda N(0, \sigma_1^2) + (1-\lambda)N(0, \sigma_2^2)$  mit  $\lambda, \sigma_1^2, \sigma_2^2$  als freien Parametern) von Interesse.

(In Übereinstimmung mit der Annahme zentraler Fehlerverteilungen sind in obigen Erörterungen die Nichtzentralitätsparameter unterdrückt. Im Zusammenhang mit stabilen Gesetzen wird das Zentrum der Verteilung im folgenden  $\delta$  genannt).

Während geschlossene Ausdrücke für die charakteristischen Funktionen der stabilen Verteilungen vorliegen, gilt dies nicht für die Dichten. Die vorliegenden Schätzprozeduren für die Parameter fußen daher entweder auf den asymptotischen Eigenschaften der Dichten oder auf der empirischen charakteristischen Funktion (Paulson, Holcomb & Leitch (1975), Koutrouvelis (1980)). Das numerisch einfache Verfahren von Koutrouvelis verlangt

1. Bildung der empirischen charakteristischen Funktion (Fourier-Transformation)  $\hat{\phi}(t)$  an mehreren Stützstellen  $t=t_k$
2. die Umformungen von  $\phi(t) = \exp(i\delta t - \gamma |t|^\alpha [1 + i\beta \operatorname{sgn}(t) \tan(\pi\alpha/2)])$ 
  - (2a)  $\log(-\log(|\phi(t)|^2)) = \log(2\gamma) + \alpha \log|t|$
  - (2b)  $\arctan(\operatorname{Im} \phi(t) / \operatorname{Re} \phi(t)) = \delta t - \beta \tan(\pi\alpha/2) \operatorname{sgn}(t) |t|^\alpha$

werden als Regressionsgleichungen ausgenutzt, wobei auf der linken Seite  $\hat{\phi}$  für  $\phi$  zu substituieren ist. Die Regressoren sind dann die Konstante und  $\log|t|$  in der ersten Gleichung,  $t$  und  $|t|^\alpha \operatorname{sgn}(t)$  in der zweiten Gleichung.
3. Rückrechnung auf  $\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}$  aus den vier Regressionskoeffizienten

Schwierigkeiten wie das Auftreten des  $\alpha$  als Teil des Regressors in (2b), welches durch den aus (2a) gewonnenen Schätzwert zu ersetzen ist oder die optimale Wahl der Stützstellen  $t_k$  sind laut Koutrouvelis durch leichte Modifikation des Verfahrens unter Kontrolle zu bringen. Man beachte jedoch, daß das Verfahren in  $\alpha = 1$  versagt. Nach den Erfahrungen des Autors liefert das Verfahren bei vielen tatsächlichen Daten  $\alpha$ -Werte nahe 2, wodurch statistisch eine Normalverteilung gewählt wird. Liegt  $\alpha$  nicht nahe 2, so kann es oft sehr

weit weg liegen, d.h. es zeigen sich keine Tendenzen zu einer Häufung um etwa 1.8, wie dies Fama (1965) in seiner klassischen Studie über Bewegungen von Börsenkursen und auch Koutrouvelis selbst im empirischen Teil seiner Arbeit erhalten.

Für die anschließende stochastische Simulation geben Chambers, Mallow & Stuck (1976) eine einfache Methode zur Erstellung stabiler Zufallszahlen an, die lediglich gleich- und exponentiellverteilte Zufallszahlen als Eingangswerte benötigt. Die Simulation sollte mit den geschätzten Parameterwerten durchgeführt werden, außer  $\alpha$  liegt "sehr nahe" bei 2. Eine genaue Testschranke für die Nullhypothese ( $H_0: \alpha=2$ ) läßt sich leider nicht angeben (siehe du Mouchel (1983)).

Das Grundproblem ist aber, mit welchen Daten in die gesamte Prozedur gegangen werden soll. (Im folgenden soll von der autoregressiven Notation  $y_t = B(L)y_{t-1} + e_t$  statt  $y_t = A(L)e_t$  ausgegangen werden) Die Residuen  $y_t - \hat{B}(L)\hat{y}_{t-1}$  sind mit AO-Ausreißern verunreinigt, die korrigierten Residuen  $r_{0,t} = \hat{y}_t - \hat{B}(L)\hat{y}_{t-1}$  hingegen sind aufgrund der beim Filterverfahren verwendeten  $\psi$ -Funktion beschränkt. Ein Vorschlag geht dahin, die aus den Originaldaten gebildeten Residuen  $r_{1,t}$  zu benutzen. Diese haben die Form

$$r_{1,t} = y_t - \hat{B}(L)y_{t-1} = y_t - B(L)y_{t-1} + B(L)y_{t-1} - \hat{B}(L)y_{t-1} - \hat{B}(L)f_{t-1} = e_t + [B(L) - \hat{B}(L)]y_{t-1} - \hat{B}(L)f_{t-1} + f_t$$

Der dabei entstehende Fehler wird im allgemeinen kleiner sein als bei

$$r_{2,t} = y_t - \hat{B}(L)\hat{y}_{t-1} = e_t + [B(L) - \hat{B}(L)]y_{t-1} + \hat{B}(L)[y_{t-1} - \hat{y}_{t-1}] + f_t$$

denn im zweiten Fall gehen die Berechnungsfehler für  $f$  voll ein, die gerade dann am stärksten ins Gewicht fallen, wenn  $e$  groß ist. Die Interpretation der  $r_{1,t}$  geht dahin, daß es sich um jene Residuen handelt, die entstehen, wenn das Vertrauen auf die Parameterschätzung des Martin-Filters groß, das Vertrauen auf die Richtigkeit der Korrektur der Daten aber gering ist.

Interessant wäre sicher auch der Vergleich der aus  $r_{1,t}$  gewonnenen Schätzwerte  $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$  mit den aus den OLS-Residuen  $r_{3,t}$  gewonnenen Schätzwerten. Der Einfluß des Terms  $[B(L) - \hat{B}(L)] y_{t-1}^*$  ist hier größer, damit der Einfluß der  $f_t$  geringer. Unterlegt man eine Mischverteilung  $\lambda I_0 + (1-\lambda)N(0, \sigma^2)$  für  $f_t$ , so ergibt sich für die charakteristische Funktion der Hauptkomponente von  $r_{1,t}$  wie  $r_{3,t}$ , nämlich  $e_t + f_t$

$$\phi_r(t) = \lambda \exp(-\gamma |t|^\alpha [1 + i\beta \operatorname{sgn}(t) \tan(\pi\alpha/2)]) + (1-\lambda) \exp(-\sigma^2 t^2 - \gamma |t|^\alpha [1 + i\beta \operatorname{sgn}(t) \tan(\pi\alpha/2)])$$

Angenommen, man schätzt  $\alpha, \beta, \gamma$  aus dieser Funktion nach dem Koutrouvelis-Verfahren, so ergibt sich tendenziell ein zu großes  $\alpha$ , ein zu kleines  $\beta$  und ein abweichendes, bei großer AO-Streuung  $\sigma^2$  eher zu großes  $\gamma$ . Der Prozeß  $(y_t^*)$  ist aber ein linearer Prozeß, der sich aus den Innovationen  $e_t$  aufbaut, unterliegt also gemäß der Grenzwertsätze einem stabilen Gesetz mit  $\alpha, \beta$  gemäß dem Gesetz der  $e_t$ . In diesem Sinne korrigiert eine schlechtere Schätzung der Transferfunktion den Bias. Die andererseits verursachte zeitliche Abhängigkeit der  $r_{3,t}$  sollte im Koutrouvelis-Verfahren keine tragende Rolle spielen.

Der Skalierungsschätzer  $\hat{\gamma}$  hingegen wird umso höher, je größer der Bias bei  $B(L)$ , da der Prozeß  $(y_t^*)$  i.a. eine höhere Skalierung aufweist als  $(e_t)$ . (Man vergleiche hier z.B. den AR(1)-Prozeß  $y_t^* = \alpha y_{t-1}^* + e_t^*$  mit  $\sigma_y^2 = \sigma_e^2 / (1 - \alpha^2)$ ).  $\gamma$  sollte also in jedem Falle aus den  $r_{1,t}$  geschätzt werden.

#### 1.4 Adaptive stochastische Prognose

Gemäß den im letzten Abschnitt vorgestellten Überlegungen sollen nun Prognosen veranstaltet werden. Im Abschnitt 1.1. haben wir festgestellt, daß die optimale univariate Prognose unter dem AO-Modell sich aus

$$y_t | t-1 = E(y_t / y_{t-1}^*, y_{t-2}^*, \dots; A(L)) = E(y_t^* / y_{t-1}^*, \dots; A(L)) = A(L) e_t$$

( $e_t := 0$ )

errechnet. Der naheliegende Schätzwert hierfür unter Anwendung des robusten Filters wäre

$$y_{t|t-1}^{(1)} = E(\hat{y}_t / \hat{y}_{t-1}, \dots; \hat{A}(L)) = \hat{A}(L) \hat{e}_t = \hat{A}(L) r_{0,t} \\ (r_{0,s} := 0, s \geq t)$$

Wie erwähnt, ist aber ein Mißtrauen gegenüber den korrigierten Daten  $\hat{y}_t$  angebracht, zumal ja der Anteil der korrigierten Werte im Verfahren (25-30%) den Anteil der tatsächlich "falschen" Daten unter der vorausgesetzten Verteilung der  $f_t$  (5-10%) bei weitem übersteigt. In diesem Sinne wäre eine Verwendung der Residuen  $r_{1,t}$  empfehlenswert. Da die  $r_{1,t}$  zur Verteilungsschätzung verwendet wurden, ist der Übergang zur stochastischen Simulation nur einleuchtend.

Für die stochastische Simulation wird  $e_t$  nicht 0 gesetzt, sondern die Information über die stochastische Struktur ausgenützt, indem in

$$y_{t|t-1}^{(2)} = \hat{A}(L) r_{1,t} \quad r_{1,s} \sim S(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, 0), s \geq t$$

zu setzen ist.  $y_{t|t-1}^{(2)}$  stellt dann ebenfalls eine Zufallszahl (theoretisch eine Zufallsgröße) dar, deren Median etwa als Punktprognose verwendet werden kann, wozu aber mehrere Realisierungen der  $r_{1,t}$  vonnöten sind. Das Stichprobenmittel sollte nicht benutzt werden, da dieses für  $\alpha < 1$  (tails der Dichte dicker als bei der Cauchyverteilung) sowieso nicht mehr konvergiert und bei  $\alpha$  etwas größer als 1 schon sehr schlechte Konvergenzeigenschaften zeitigt. Die Anzahl der Simulationen ist natürlich eine Frage der Rechenzeit, das Erreichen von Präzisionsschranken ist von den Parametern  $\alpha, \beta, \gamma$  abhängig. Fraktildaten für die stabilen Verteilungen könnten hier hilfreich sein, haben aber bis dato noch keine weite Verwendung gefunden.

Verfügt man über ein multivariates Zeitreihenmodell, das etwa mithilfe von Kausalitätstests schon hinreichend nach seiner Struktur abgetestet ist, so ergibt sich die beste Prognose unter einem Vektor-AO-Modell aus

$$y_{t|t-1,i} = E(y_{t,i} / y_{t-1}^*, y_{t-2}^*, \dots)$$

wobei  $y_t, y_t^*$  Vektoren seien mit  $y_{t,i}, y_{t,i}^*$  etc. als i-ter Komponente. Grundsätzlich ist das robuste Filterverfahren auf multivariate Prozesse komponentenweise anwendbar, Erfahrungswerte liegen hier leider nicht vor. Die Nichtberücksichtigung der instantanen Korrelationen zwischen den Komponenten könnte zu einer

wesentlichen Ineffizienz führen, auch die Reihung der Komponenten könnte eine Rolle spielen. Die Anzahl der Iterationen wächst sicher mit steigender Dimension stark an.

Ausgehend davon, daß die meisten Rechenanlagen über robuste multivariate Zeitreihenschätzer nicht verfügen, scheint es taktisch optimal, komponentenweise Regressionen mit herkömmlichen M- oder GM-Prozeduren durchzuführen, um die Transfermatrix  $A(L)$  bestimmen zu können. Da hier keine Datenbereinigung stattfindet, können die entstehenden Residuen sofort auf ihre Verteilung hin untersucht werden, wobei im Auge zu behalten ist, daß die Schätzungen gemäß der in 1.3. angestellten Überlegungen verzerrt sind. Der sich aufdrängenden stochastischen Simulation unter

$$r_{t,i} \sim S(\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i, 0)$$

stellen sich also zusammenfassend folgende Probleme in den Weg:

1. Die gemeinsame Verteilung der  $(r_{t;1}, \dots, r_{t;n})$  bleibt unberücksichtigt. Mit Ausnahme des Falls  $\hat{\alpha}_1 = \dots = \hat{\alpha}_n = 2, \hat{\beta}_1 = \dots = \hat{\beta}_n = 0$ , also des rein normalverteilten Modells, scheint es unmöglich, auf diese näher einzugehen, da mehrdimensionale stabile Verteilungen enorme Probleme aufwerfen (siehe Press(1972)).
2.  $\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i$  sind verzerrte Schätzer für  $\alpha_i, \beta_i, \gamma_i$  aufgrund der AO-Einflüsse. Die geschätzte Verteilung tendiert zu "guten" Verteilungen mit zu "schlechter" Skalierung. Welcher Effekt überwiegt, ist schwer zu sagen.
3. Die zur Prognose benützten Werte der Vorperioden zum Prognosezeitraum sind durch AO-Einflüsse verfälscht. Entsteht hier eine zufällige Häufung (patchy outliers), so gerät die Prognose kurz- bis mittelfristig außer Kontrolle.
4. Je stärker die gegenseitige Abhängigkeit der Komponenten  $y_{t,i}$  ist, desto stärker tendieren die  $\alpha_i$  der Verteilungen der  $y_{t,i}$  zum  $\min(\alpha_i)$ . Starke Abweichungen zwischen A und A verzerren dann die  $\alpha_i$ -Schätzung nach unten.

## II. Durchführung

### 2.1. Daten

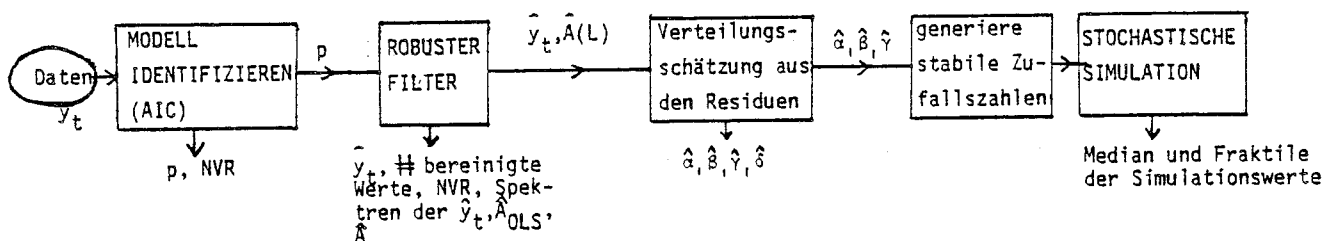
Anlaß zur Erstellung der vorliegenden Arbeit waren die bei einem Zeitreihenmodell für die österreichische Wirtschaft (Kunst (1985)) aufgetretenen Ausreißerprobleme. Als Grundlage für eine Untersuchung bieten sich an:

- die einzelnen 12 Variablen des Zeitreihenmodells in univariaten Modellen
- das multivariate Gesamtmodell unter Beibehaltung der (ohne Zuhilfenahme robuster Methoden) identifizierten Struktur
- verwandte Modelle, univariate Modelle anderer Variablen, vor allem exogene Größen in üblichen makroökonomischen Modellen

Inwiefern eine sicherlich aufwendige robuste Neuschätzung des Modells in b, d.h. der zugrundeliegenden Kausalitätstests, wünschenswert ist, ist noch zu klären.

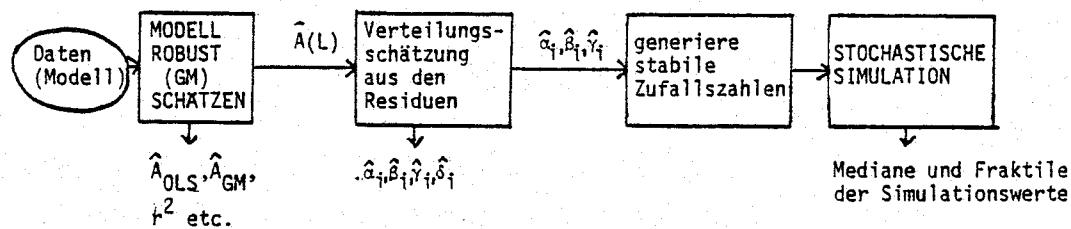
### 2.2. Software

Für die univariaten Modelle ergibt sich, aufbauend auf Abschnitt I, folgendes Flußdiagramm der Durchführung:



Stufe 1 und 5 können mit dem am Institut für Höhere Studien zur Verfügung stehenden IAS-System durchgeführt werden, für die Stufen 2-4 (Stufe 2 nur für das AR-Modell) existieren vom Autor geschriebene oder im Zuge der Erstellung befindliche Fortran-Programme (für Stufe 2 in älteren Version in Bomze et al.(1983) dokumentiert). In der unteren Zeile sind jene Werte angegeben, die von Interesse bei einer Dokumentation sein könnten.

Für das multivariate Modell ergibt sich ein modifiziertes Flußdiagramm



Für Schritt 1 ist derzeit noch keine Prozedur auf der Rechenanlage verfügbar. Ebenso wenig existiert eine multivariate Version des robusten Filters.

### 2.3 Erste empirische Evidenz: robuste Spektren

Als ein erster Schritt wurden die 12 in Kunst(1985) verwendeten Variablen, also Quartalswerte für

- CD ... dauerhafter Konsum, real
- CNDS ... nicht dauerhafter Konsum, real
- IFE ... Ausrüstungsinvestitionen, real
- IFC ... Bauinvestitionen, real
- M01\$ ... Importe landwirtschaftlicher Produkte nominell
- X01\$ ... Exporte landwirtschaftlicher Produkte nominell
- M24\$ ... Importe von Rohstoffen nominell
- X24\$ ... Exporte von Rohstoffen nominell

M59\$ ... Importe von Industriegütern (SITC 5-9) nominell  
X59\$ ... Exporte von Industriegütern (SITC 5-9) nominell  
PGDP ... BIP-Deflator  
RSECL .. Rendite der Neuemissionen langfristiger Bonds

univariat autoregressiv modelliert. Da hier auf eine Beschränkung der AR-Ordnung  $p$  verzichtet werden konnte, sind die folgenden Resultate mit denen der obzitierten Arbeit (siehe Tafel 2 dort) nicht direkt vergleichbar. Überdies wurde der Erhebungszeitraum bis 1984.4 ausgedehnt. Unter Vorgabe des identifizierten  $p$  (gemäß AIC) wurde nun der robuste Filter angewendet. Die Statistiken für unbereinigte und bereinigte Werte sind in Tabelle 1 zusammengefaßt. Im Anhang präsentieren sich Schaubilder nichtparametrischer Spektren der bereinigten wie unbereinigten Daten, wobei, der Samplegröße von etwa 80-120 entsprechend, Fensterweiten von 20/30 bzw. 15/25 für die Spektralschätzung unterlegt wurden. Das Parzen-Spektralfenster gilt als Standard des IAS-Systems. (Die auf den Ausdrucken erscheinenden Erweiterungen der Labels um die Buchstaben D, L oder L4, R oder eine Ziffer beziehen sich auf die gebrauchten Transformation der aus der WIFO-Datenbank übernommenen Daten (D=1. Differenzen, L=L4=1. und 4. Differenzen nach Logarithmieren) bzw. auf die Filterung).

In der Tabelle sind angegeben: das Label der Zeitreihe; der NVR (noise/variance-ratio) des mit AIC-Ordnung  $p$  identifizierten Originalmodells; der NVR des AR( $p$ )-Modells der robust gefilterten Daten; der NVR des mit AIC-Ordnung  $p'$  identifizierten Modells der gefilterten Daten, wo  $p'=p$ ;  $p$  bzw.  $p'$ ; die Anzahl der Daten; die Anzahl der hievon korrigierten Daten ( $f_t=0$ ); der Zeitpunkt der ersten Beobachtung (letzte ist immer 1984.4); die Anzahl der Iterationen des robusten Filters bis zur Konvergenz.

Wo sich  $p'$  von  $p$  unterschied, wurde mit  $p'$  nochmals in den robusten Filter gegangen und solange iteriert, bis sich AIC- $p$  ex ante und AIC- $p$  ex post nicht mehr unterschieden. Die Ergebnisse dieser Prozedur sind in den zweiten Zeilen unter der jeweiligen Variable angegeben. Die Spektren beziehen sich auf diese Daten bzw. Modelle mit oft viel höherem  $p$  als zu Beginn.



Label	NVR <sub>0</sub>	NVRP <sub>p</sub>	NVRP <sub>p'</sub>	p	p'	Daten	korr.	1.Beob.	Iter.
CD	.564	.485	.440	8	10	119	35	1955.2	5
		.393		10			39		6
CNDS	.460	.390		8		119	42	1955.2	5
IFE	.393	.391		16		79	21	1965.2	7
IFC	.922	.946		1		79	29	1965.2	3
MØ1	.575	.465	.478	16	14	127	37	1953.2	4
XØ1	.601	.525		11		127	41	1953.2	4
M24	.579	.602		13		127	36	1953.2	5
X24	.543	.557		12		119	36	1955.2	5
M59	.600	.603		16		127	30	1953.2	5
X59	.511	.451		12		119	35	1955.2	5
PGDP	.727	.658	.640	4	5	79	20	1965.2	4
		.647		5			21		5
RSECL	.913	.578	.524	3	7	124	46	1954.1	10

## 2.4 Versuch einer Interpretation der Spektren

Nahezu alle Diagramme widerspiegeln den Effekt der Vorbereitung der Daten durch erste und vierte Differenzen durch Einbrüche bei  $0, \pi/2, \pi$  und verstärkte Frequenzen etwa in der Mitte zwischen diesen Werten.

Die Außenhandelsdaten zeigen ein recht homogenes Bild. Die durch die Transformation hervorgerufenen Muster dominieren, deutliche Spitzen bei anderen Frequenzen fehlen. Bei den robusten Spektren entspricht allerdings die Auflösung bei der geringeren Fensterweite bereits der Auflösung bei höherer Fensterweite in Bezug auf die Originaldaten. Im einzelnen gilt für die Importe und Exporte landwirtschaftlicher Produkte (M01 bzw. X01), daß die starken Zyklen (8 bzw. 9 Monate und in beiden Fällen etwa 2 1/2 Jahre) bereits bei den nicht korrigierten Reihen deutlich erkennbar sind, die Ausreißerbereinigung also keinen Informationsgewinn bringt. Man vergleiche die schnelle Konvergenz des robusten Filterverfahrens. Bei den Rohstoffimporten (M24) zeigt sich sogar eine Verflachung der bei den Originaldaten ausgewiesenen Spitzen. Lediglich die Rohstoff- und Energie-Exporte (X24) lassen deutlichere Schwingungen erkennen (8 und 16 Monate sowie 3 Jahre), deren Auflösung durch Filterung verbessert wird.

Bei den Exporten an Industriegütern (X59) verbessert die Filterung die Auflösung im kurzwelligen Bereich, neben eine 7-Monats-Welle tritt nun eine 9-Monats-Welle, während sich der langfristige Zyklus nach links verschiebt, eine Länge von etwa 3 Jahren erreichend, in Übereinstimmung mit der Außenhandelsentwicklung bei den Rohstoffen. Importseitig ergeben sich bei den gefilterten Daten noch stärkere Verschiebungen, das Spektrum von M59\$LR läßt Zyklen von 7 und 9 Monaten sowie 2 und 4 Jahren erkennen, die in der Standardversion schlechter auffindbar sind. Eine ökonomische Interpretation könnte dahin gehen, daß kurzfristig (7 und 9 Monate) Exporte und Importe parallel laufen, langfristig aber verschiedene andere (Konjunktur- ?) Elemente zum Tragen kommen. Summa summarum wirken die Außenhandelsdaten "over-differenced", was allerdings bereits bekannt war und im Hinblick auf die analoge Behandlung der übrigen Zeitreihen hingenommen wurde.

Der nichtdauerhafte Konsum (CNDS) zeigt im Original ein flaches Spektrum mit viel Masse im kurzwelligen Bereich. Erst nach Filterung lassen sich die dominanten Zyklen ( $2 \frac{1}{4}$  Jahre und 17 Monate) im langwelligen Bereich erkennen. Auch im kurzwelligen Bereich trennt sich der "Tafelberg" in zwei kleine Gipfel. Im wesentlichen erscheint der nicht-dauerhafte Konsum als von langen Konjunkturwellen unabhängig. Das schönste Beispiel für die Ausreißerberinigung überhaupt stellt hingegen der dauerhafte Konsum (CD) dar, der in der bereinigten Version schon bei kleiner Fensterlänge die Zyklen auflöst, die die unbereinigten Daten verschweigen: eine lange Welle von mehr als fünfjähriger und eine mittlere von eineinhalbjähriger Dauer, welche die Lebensdauer respektive Produktinnovation von langfristigen (z.B. technischen) und mittelfristigen (z.B. textilen) Konsumgütern recht gut widerspiegeln könnte. Die relative Masse im kurzwelligen Bereich ist auch bei CD hoch, allerdings naturgemäß nicht so stark wie bei CNDS. Der eklatante Gewinn an Struktur bei Ausreißerbereinigung bezüglich CD geht mit den Schwierigkeiten konform, die bei Kunst(1985) aufgetreten sind und vor allem durch Mehrwertsteuerumstellungen hervorgerufen wurden. Das Filterverfahren scheint imstande, diese Effekte weitgehend zu isolieren.

Ähnlich wie der dauerhafte Konsum reagieren auch die Ausrüstungsinvestitionen (IFE) positiv auf die robuste Filterung. Die Länge der Investitionszyklen ( $3 \frac{1}{2}$  und  $1 \frac{1}{2}$  Jahre) läßt sich aus den Originaldaten nicht ablesen. Wie schon beim dauerhaften Konsum, ist dieses Resultat in Übereinstimmung mit der relativ großen Anzahl an Iterationen des Filterverfahrens, die, besser als der Anteil der korrigierten Daten, auf eine hohe Ausreißerbelastung hinweist. Die Interpretation von Ausreißern im Investitionsverhalten ist allerdings nicht so einfach. Die Bauinvestitionen (IFC) sind in diesem Zusammenhang eigentlich nur einer Erwähnung am Rande wert. Sie entsprechen einem negativ korrelierten AR(1)-Modell, alle erkennbaren Abweichungen der Spektren wirken da rein zufällig.

Die höchste Anzahl an Iterationen (10) ist beim langfristigen Zinssatz (RSECL) aufgetreten. Die Umwandlung von 38% der Daten reflektiert auch das Spektrum. Zeigen die Originaldaten einen langen Zyklus von  $4 \frac{1}{2}$  Jahren und einen kurzen Zyklus von  $9 \frac{1}{2}$  Monaten neben einer schwachen Halbjahreswelle, verändert sich das Bild nach Filterung völlig. Der langwellige Zyklus bleibt aufrecht, nähert sich aber eher 5 Jahren an, während der kurzwellige Teil zu flachem "noise"

degeneriert. Auch der Übergang von gebrochen ganzzahligen Zinssatzdaten zu reellen Daten, der automatisch aus der Filterung folgt, kann hier eine Rolle spielen. Die lange Welle des Zinsgefüges stimmt von der Dauer her am ehesten mit derjenigen des dauerhaften Konsums überein.

Der einzige Preisindex auf der Liste, der GDP-Deflator, verläßt die Filterprozedur ohne erkennbare Änderungen. Zwei sanfte Wellen bei 9 und 20 Monaten treten auf, entsprechen aber leider fast der Frequenzdarstellung der Transformation und sind so ohne große Aussagekraft.

## 2.5 Verteilungsschätzung anhand der Rohdaten

In Abschnitt 1.3. wurde auf die Parameterschätzung auf der Basis stabiler Verteilungsgesetze nach Koutrouvelis eingegangen. Als ein erster Test wurden nun anstatt der Residuen eines Zeitreihenmodells zunächst die Rohdaten diesem Verfahren unterzogen. Zum Vergleich und als besonderer Hinweis auf mögliche Defekte nichtstochastischer Prognose wurde anschließend das Schätzverfahren bis 1990 ausgedehnt, wobei die Prognosedaten aus der jüngsten Version des in Kunst(1985) geschilderten multivariaten Zeitreihenmodells zu den Beobachtungen hinzugefügt wurden. Die sich ergebenden Schätzwerte für  $\alpha, \beta, \gamma, \delta$  sind in der Tabelle 2 angegeben. (in Klammern der erweiterte Zeitbereich)

Viele der Zeitreihen erwecken also den Verdacht auf Abweichungen von der Normalverteilungsannahme. Die Anzahl der Stützstellen für die beiden Koutrouvelis-Regressionen (K,L) wurde auf einen, dem optimalen bei Koutrouvelis angegebenen nahestehenden Wert fixiert, eine bessere Approximation scheint aufgrund des geringen Samples unmöglich (die optimalen K,L-Werte bei Koutrouvelis beziehen sich auf mindestens 200 Daten). Die Parameterschätzer reagieren allerdings auf eine Änderung von K,L nicht sehr sensibel. Die größten Schwierigkeiten bei der Schätzung macht  $\beta$  und damit auch  $\delta$ , die beide aus dem Phasenwinkel der charakteristischen Funktion bestimmt werden. Speziell für  $\alpha$  nahe 2 ergeben sich dabei oft schlechte Werte, was auch darauf zurückzuführen ist, daß dann der Wert  $\tan(\pi\alpha/2)$  numerisch nahe 0 liegt. Selbst  $\beta$ -Werte außerhalb des zulässigen Intervalls  $(-1,1)$  treten auf.

label	$\alpha$	$\beta$	$\gamma$	$\delta$	K=L
CD	1.37 (1.14)	-.19 (-.54)	.014 (.021)	.0076 (.0697)	15
CNDS	1.88 (1.79)	1.02 (.65)	.0005 (.0005)	-.0015 (-.0012)	10
IFE	1.58 (1.23)	-.18 (.12)	.0077 (.0138)	.0022 (-.0105)	15
IFC	1.53 (1.20)	-.12 (.24)	.0024 (.0055)	.0001 (-.0091)	15
MØ1	1.76 (1.62)	.28 (-.03)	.0139 (.0143)	.0001 (.0016)	10
XØ1	1.42 (1.24)	.05 (-.35)	.057 (.058)	-.0040 (.0771)	15
M24	1.50 (1.43)	-.13 (-.06)	.0122 (.0118)	.0081 (.0042)	15
X24	1.76 (1.68)	.42 (.44)	.0068 (.0069)	-.0054 (-.0069)	12
M59	1.69 (1.57)	-.96 (-.69)	.0060 (.0065)	.0116 (.0109)	12
X59	1.74 (1.68)	.26 (.42)	.0026 (.0025)	-.0029 (-.0042)	10
PGDP	1.69 (1.41)	-.68 (.02)	.0004 (.0009)	.0014 (-.0009)	15
RSECL	1.23 (.85)	.54 (-.14)	.088 (.125)	-.1522 (-.0445)	15

Deutlich ist zu erkennen, daß die nichtstochastische Prognose den Wert für  $\alpha$  drückt, da nun ein Übermaß an Werten nahe dem Verteilungszentrum liegt und die Prozedur ein leptokurtisches Gesetz identifiziert. Weiters verschieben sich Zentrum und Schiefe, was auf den Einfluß der übrigen Variablen im multivariaten Modell zurückzuführen sein dürfte. Als Beispiel möge der Zinssatz (RSECL) dienen, dessen deutliche positive Schiefe (langsame Zunahme, rapides Abbröckeln) von der Prognose überhaupt nicht wiedergegeben wird.

## LITERATUR

BOMZE, I.M., R.M. KUNST & E. RESCHENHOFER(1983): Über die Schätzung von Spektraldichten; Forschungsbericht 184, IHS Wien.

CHAMBERS, J.M., C.L. MALLOWS & B.W. STUCK(1976): A New Method for Simulating Stable Random Variables; JASA 71,340-344.

COOK, R.D.(1977): Detection of influential observations in linear regression; Technometrics 19,15-18.

DUMOUCHEL, W.H.(1983): Estimating the Stable Index in Order to Measure Tail Thickness: A Critique; Annals of Statistics 11,1019-1031.

FAMA, E.F.(1965): Behavior of Stock Market Prices; Journal of Business 38,34-105.

FOX, A.J.(1972): Outliers in Time Series; JRSS B 34,350-363.

HANNAN, E.J. & M. KANTER(1977): Autoregressive Processes with Infinite Variance; Journal of Applied Probability 14,411-415.

KLEINER, B., R.D. MARTIN & D.J. THOMSON(1979): Robust Estimation of Power Spectra; JRSS B 41,313-351.

KOUTROUVELIS, I.A.(1980): Regression-Type Estimation of the Parameters of Stable Laws; JASA 75,918-928.

KUNST, R.M.(1984): (Dissertation)

KUNST, R.M.(1985): Ein Zeitreihenmodell für die österreichische Wirtschaft; Forschungsbericht 211, IHS Wien.

LEDOLTER, J.(1985): Leverage, Influence and Outliers in Time Series Models; Working Paper, University of Iowa.

MARONNA, BUSTOS & YOHAI (1979): Bias in Efficiency Robustness of General M-estimates with Random Carriers; in: Th.GASSER & M.ROSENBLATT(ed.): Smoothing Techniques for Curve Estimation, Springer, New York.

- MARTIN, R.D.(1979): Approximate Conditional Mean Type Smoothers and Interpolators; in: ob.cit.
- MARTIN, R.D.(1981): Robust Methods for Time Series; in: D.F.FINDLEY(ed.): Applied Time Series Analysis II, Academic Press, New York.
- MARTIN, R.D., A.SAMAROV & W.VANDAELE(1981): Robust Methods for ARIMA Models; Technical Report 29, MIT.
- PAULSON, A.S., E.W. HOLCOMB & R.A. LEITCH(1975): The Estimation of the Parameters of the Stable Laws; Biometrika 62,163-170.
- PRESS, S.J.(1972): Estimation in Univariate and Multivariate Stable Distributions; JASA 67,842-846.
- THOMSON, D.J.(1977): Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide-I; Bell Systems Technical Journal 56,1769-1815.
- YOHAI, V.J. & R.A. MARONNA(1977): Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variances; Annals of Statistics 5,554-560.



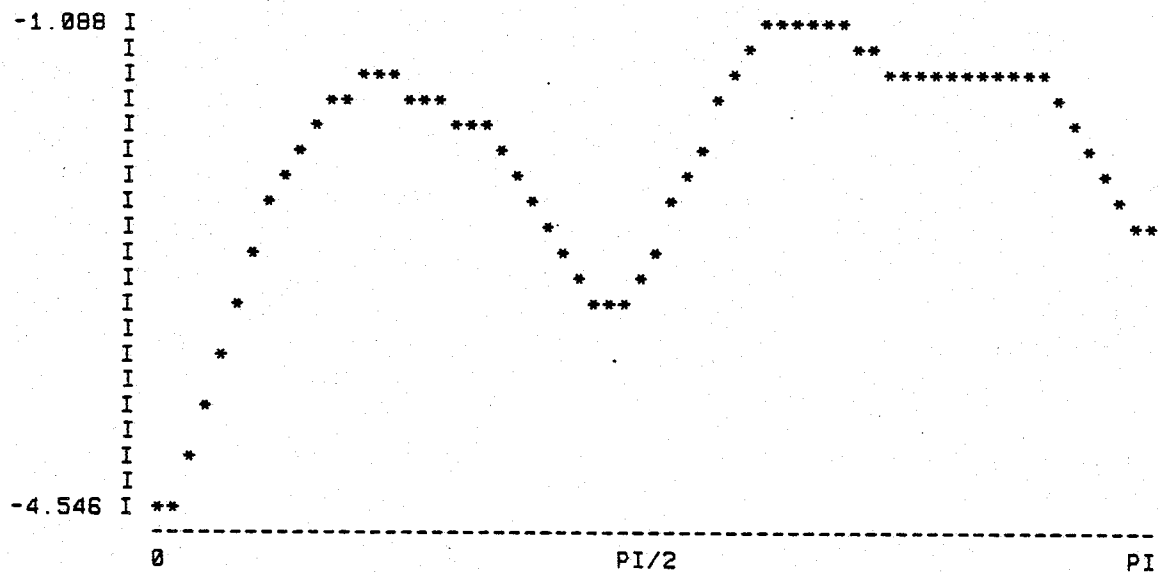


ANHANG



\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M01\$L4

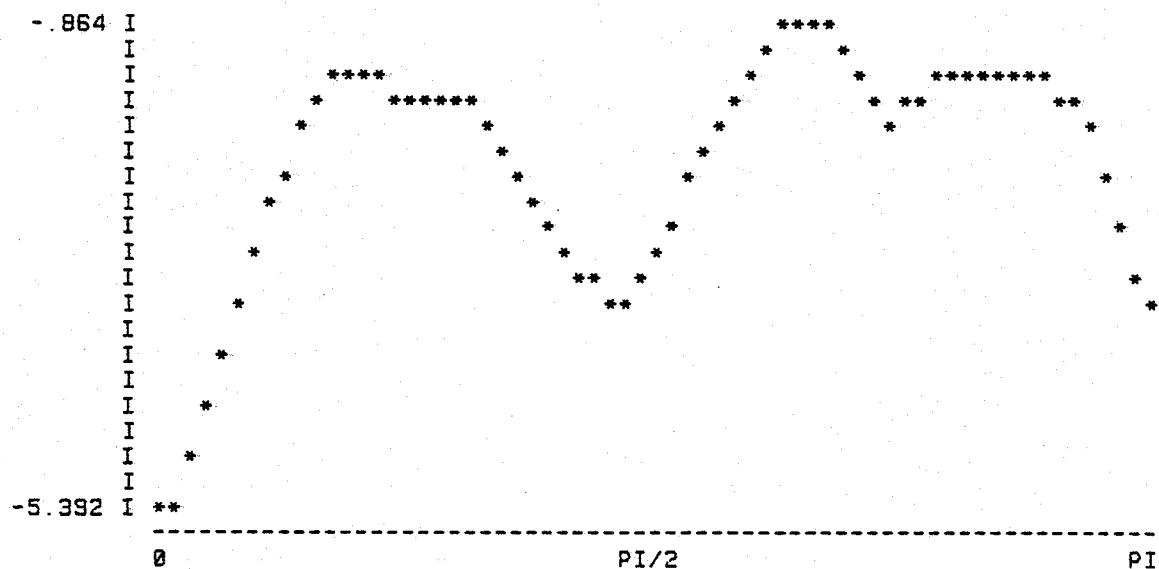


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M01\$L4

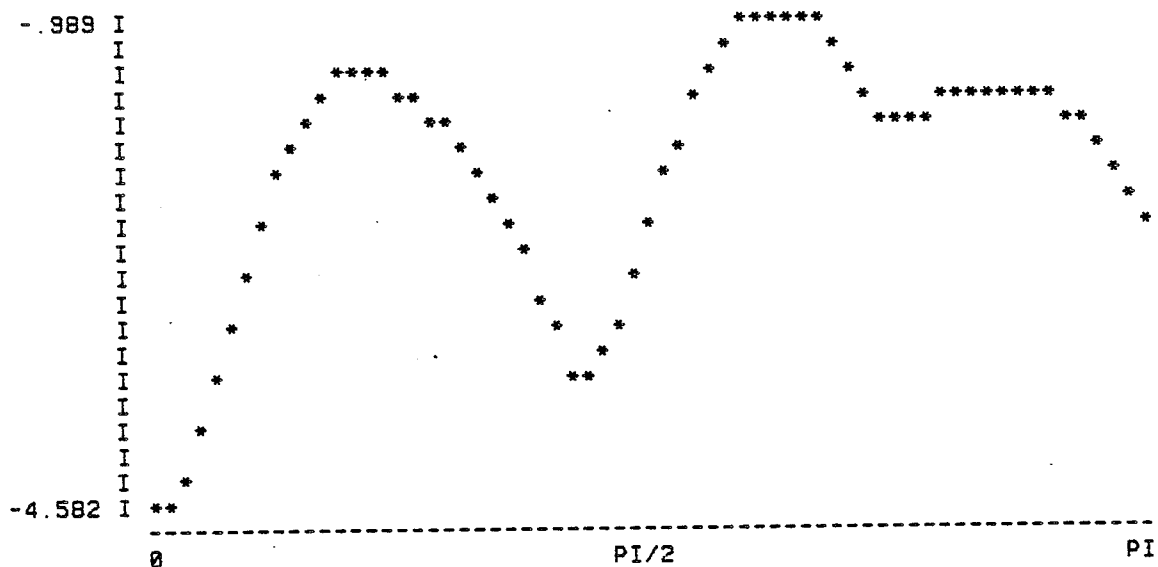


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M01\$LR

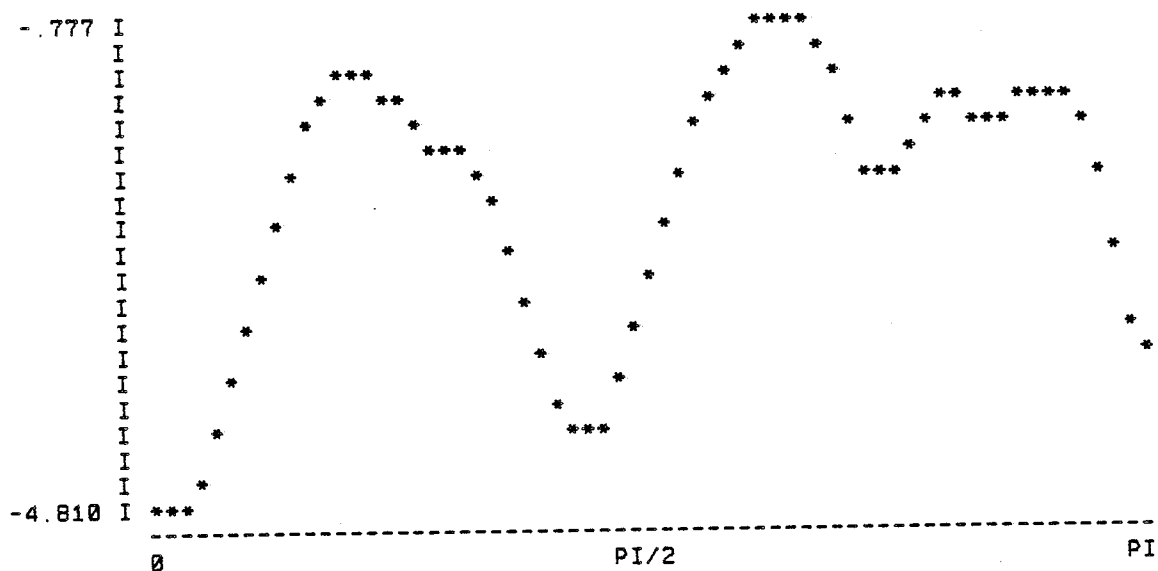


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M01\$LR

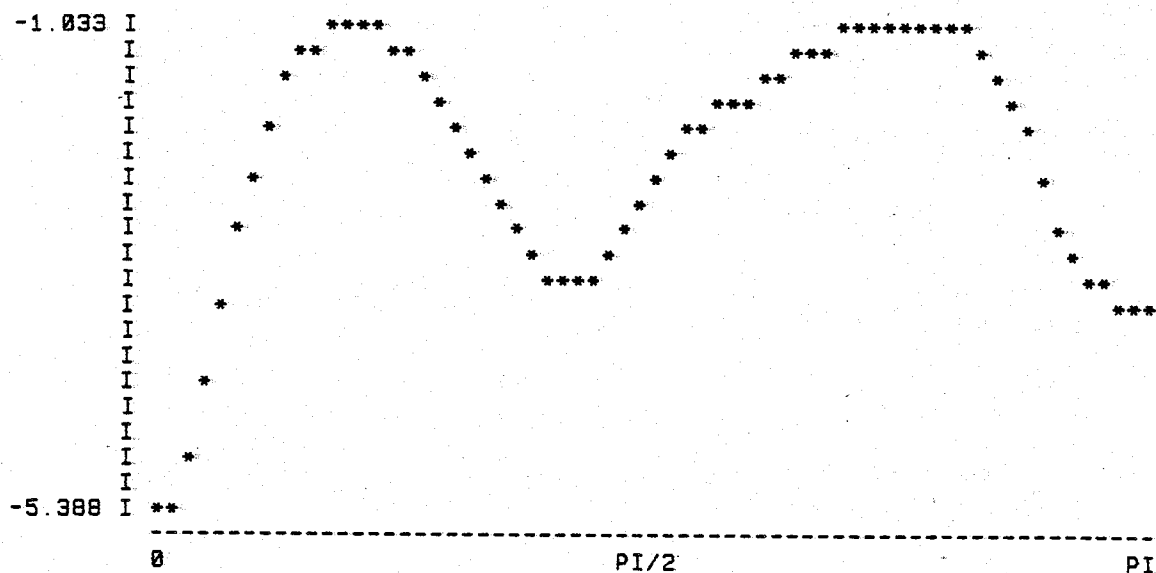


Parzen Window of Length : 30

\*\*\*\*\*

A scatter plot showing the function  $f(x) = 1 - \cos(2x)$  for  $x$  in  $[0, \pi]$ . The x-axis is labeled with  $0$ ,  $\pi/2$ , and  $\pi$ . The y-axis is labeled with  $-4.336$  and  $-1.122$ . The plot shows a periodic wave starting at  $(0, 0)$ , reaching a minimum at  $(\pi/2, -4.336)$ , and returning to  $0$  at  $(\pi, 0)$ .

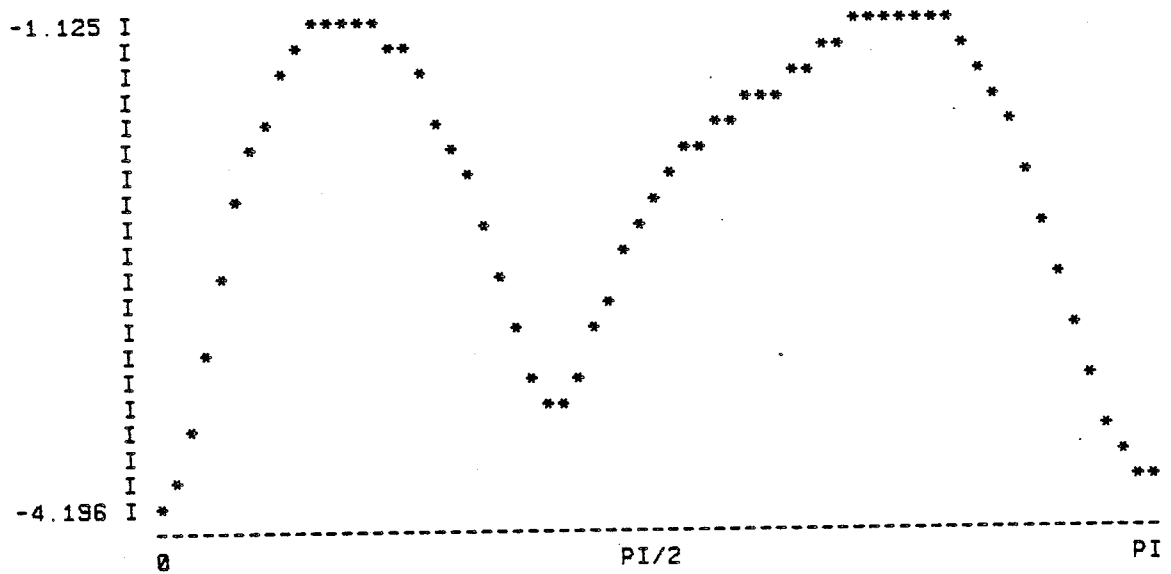
LOGGED NORMALIZED SPECTRUM OF : X01\$L4



Parzen Window of Length : 30

\*\*\*\*\*

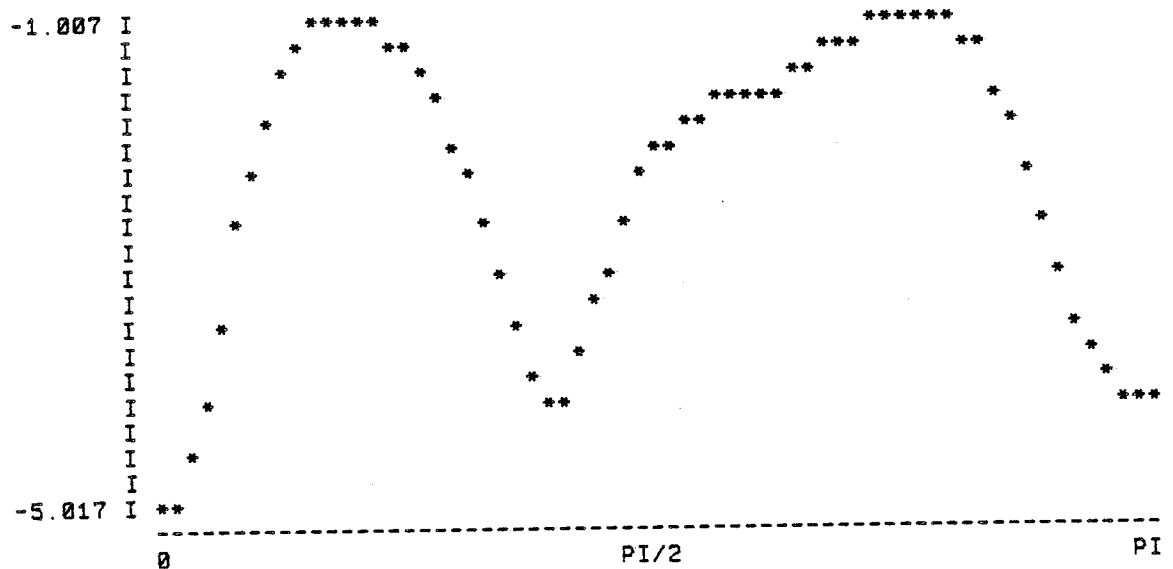
LOGGED NORMALIZED SPECTRUM OF : X01\$LR



\*\*\*\*\*

\*\*\*\*\*

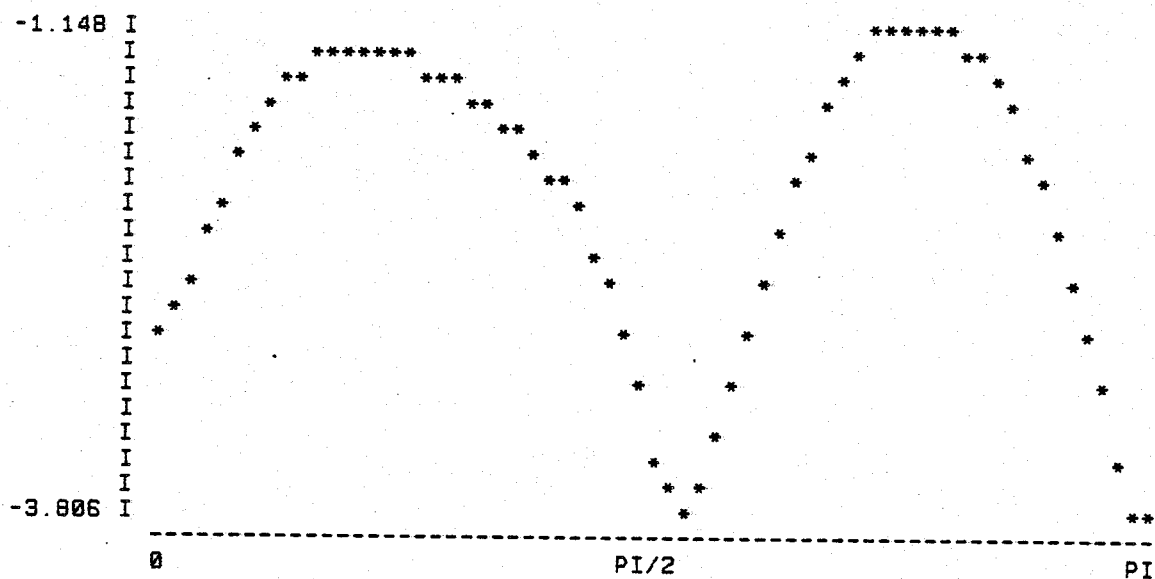
LOGGED NORMALIZED SPECTRUM OF : X01\$LR



\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M24\$L4

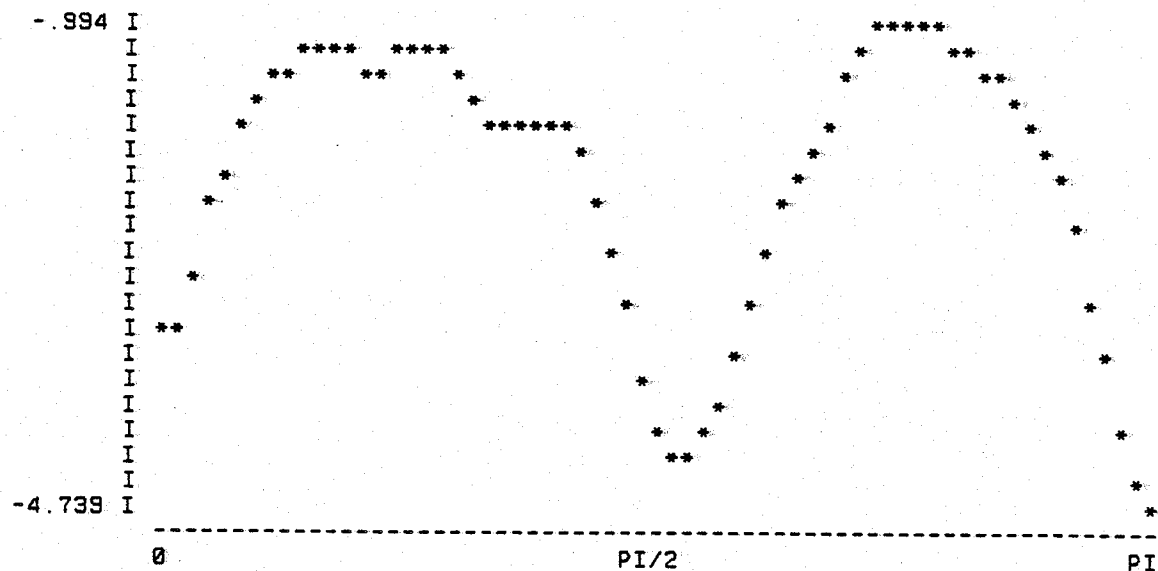


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M24\$L4

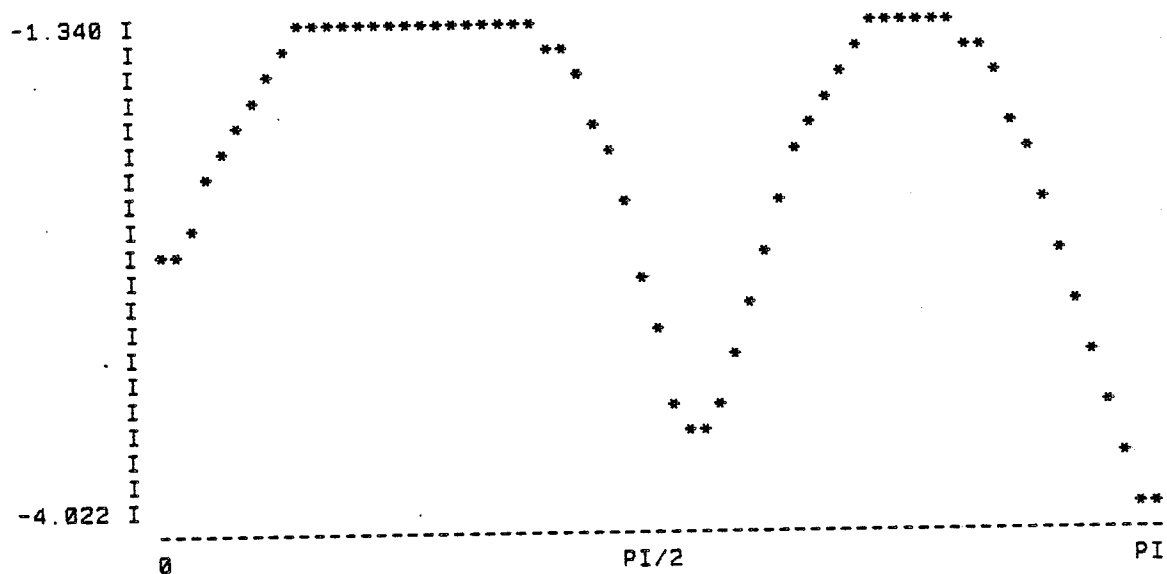


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

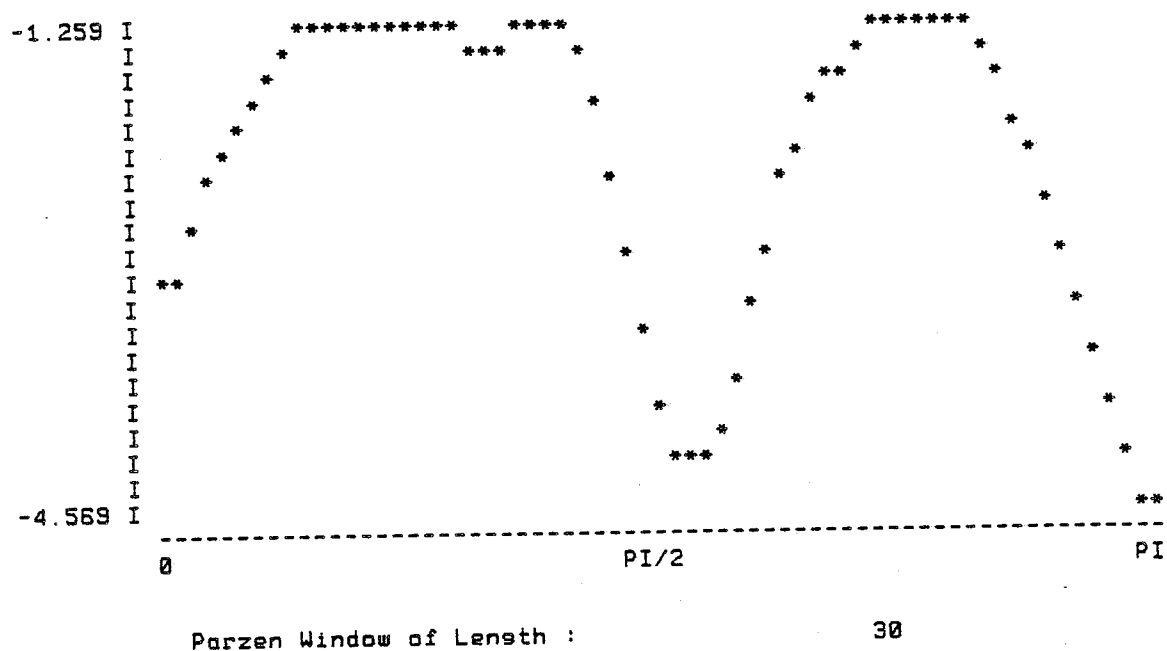
LOGGED NORMALIZED SPECTRUM OF : M24\$LR



\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : M24\$LR



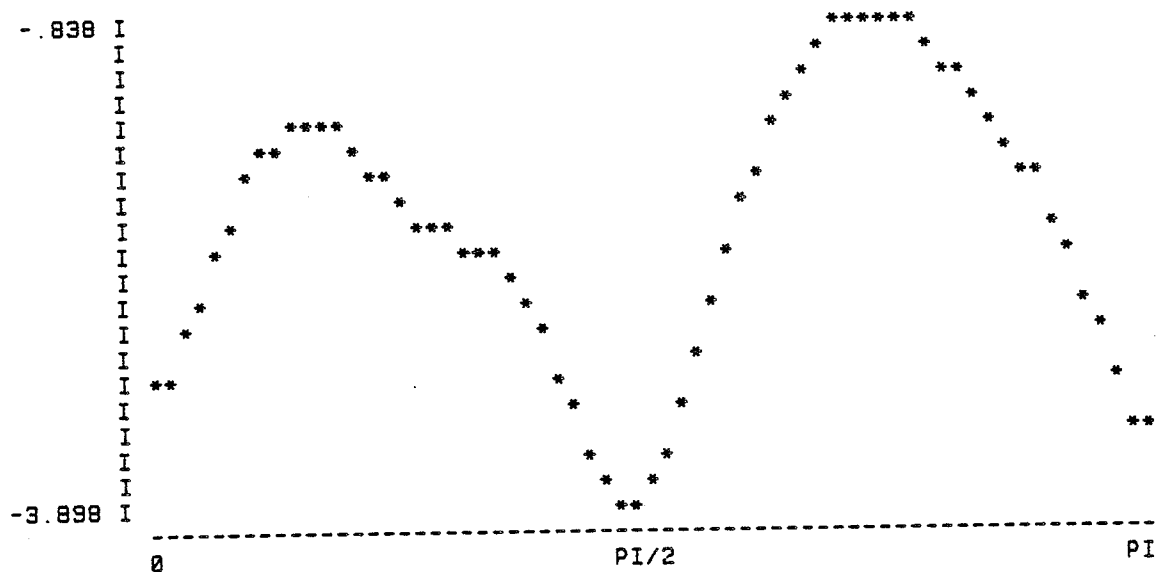
\*\*\*\*\*





\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X24\$LR

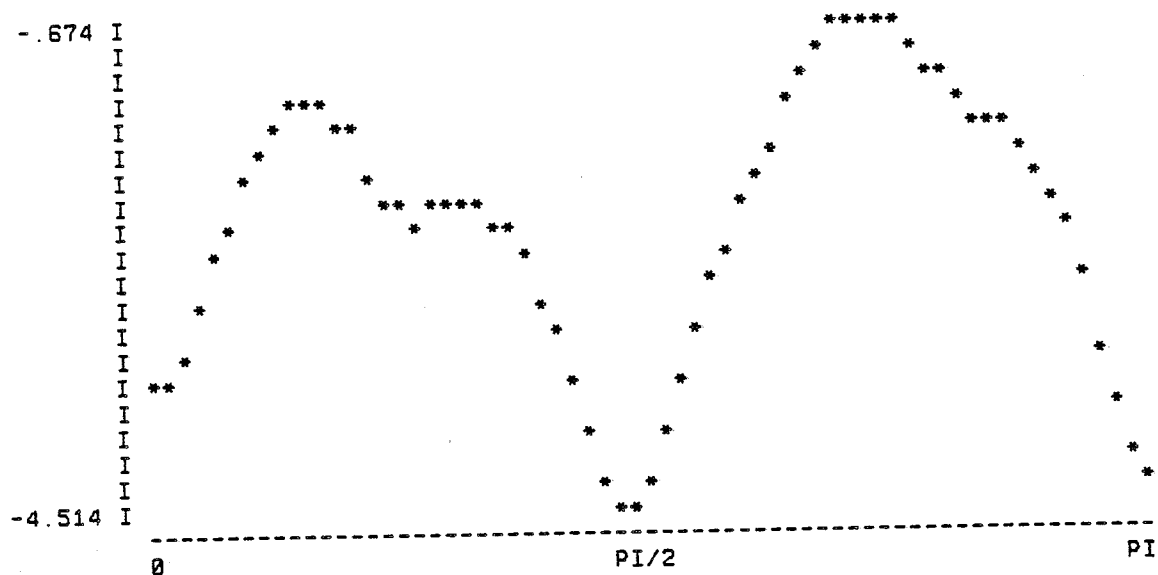


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X24\$LR

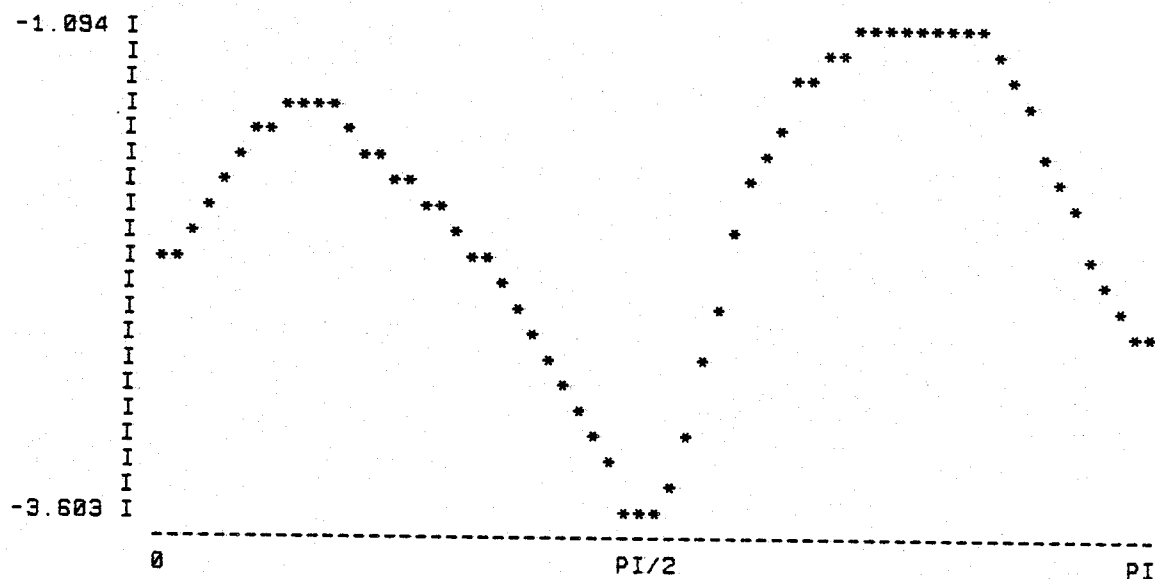


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

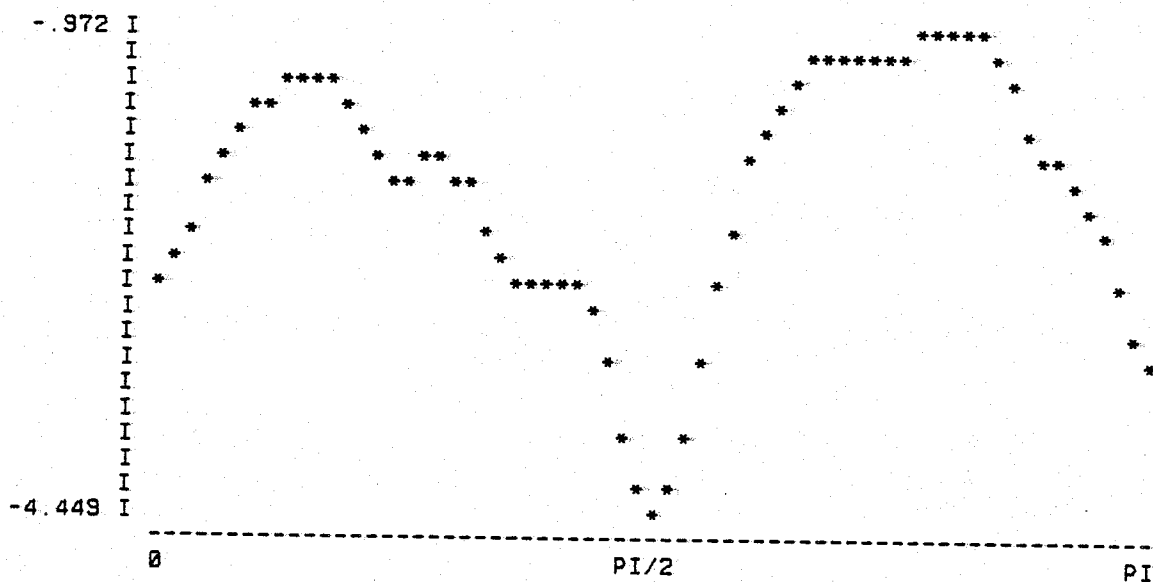
LOGGED NORMALIZED SPECTRUM OF : M59\$L4



\*\*\*\*\*

\*\*\*\*\*

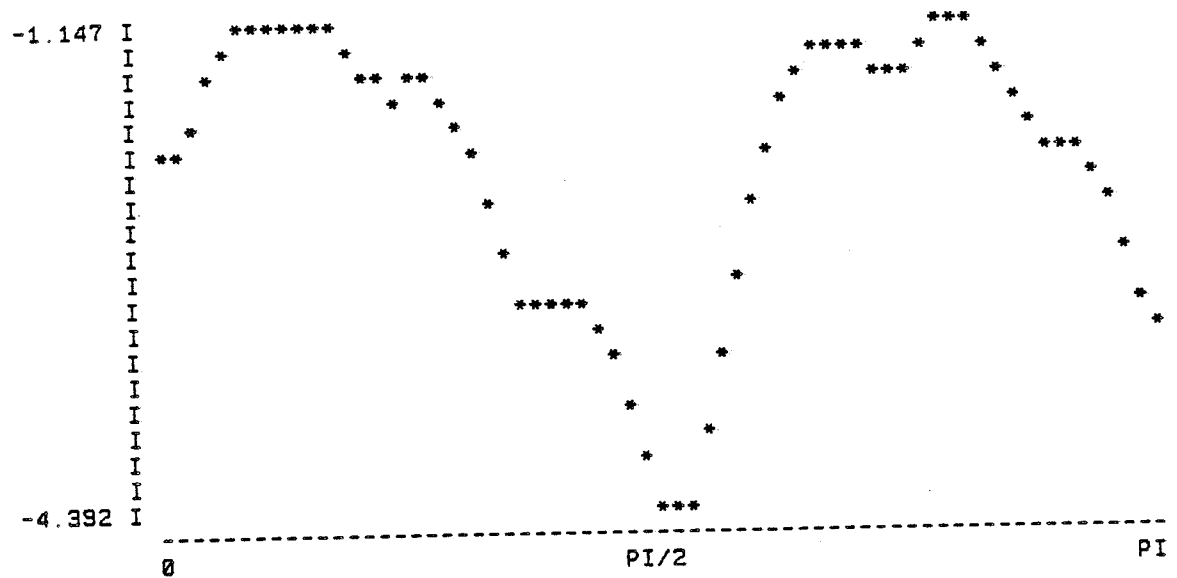
LOGGED NORMALIZED SPECTRUM OF : M59\$L4



\*\*\*\*\*

The figure displays a scatter plot of the function  $f(x) = 1 - 2\cos(2x)$  for  $x \in [0, \pi]$ . The x-axis is labeled with  $0$ ,  $\pi/2$ , and  $\pi$ . The y-axis has labels at  $-1.231$  and  $-3.766$ . The data points, represented by asterisks, form a periodic wave that starts at  $(0, -1.231)$ , reaches a minimum of  $-3.766$  at  $x = \pi/2$ , and returns to  $-1.231$  at  $x = \pi$ . A dashed horizontal line is drawn at  $y = -3.766$ .

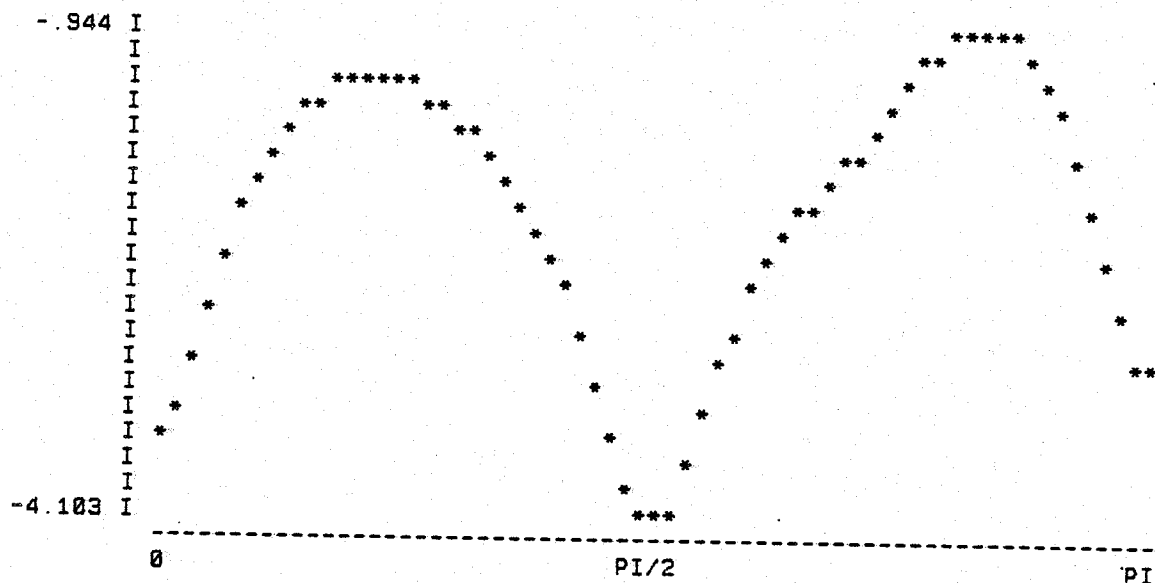
LOGGED NORMALIZED SPECTRUM OF : M59\$LR



Parzen Window of Length : 30

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X59\$L4

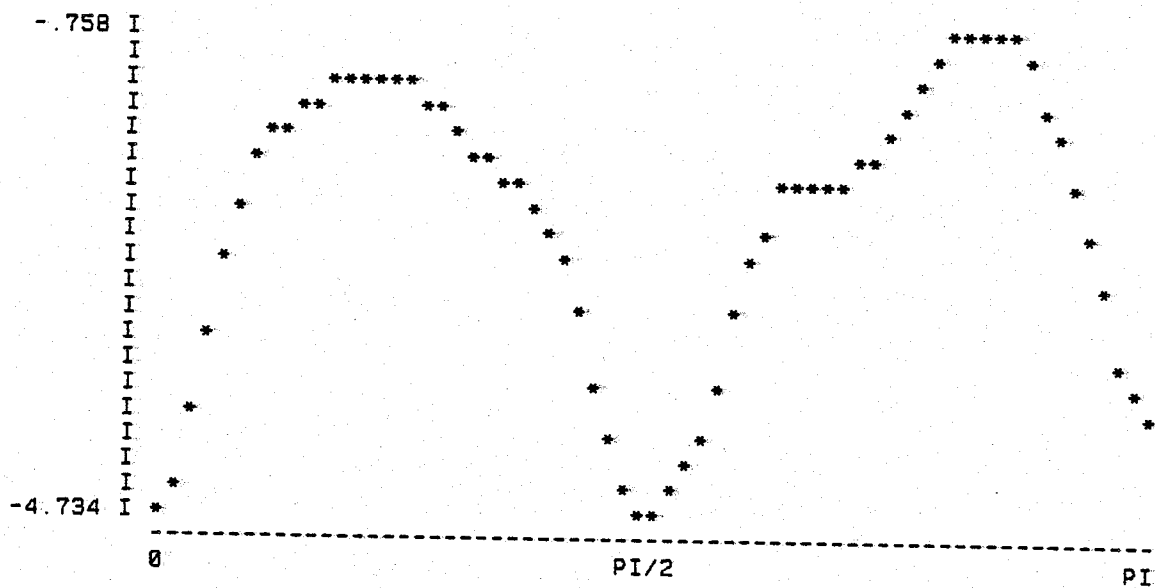


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X59\$L4

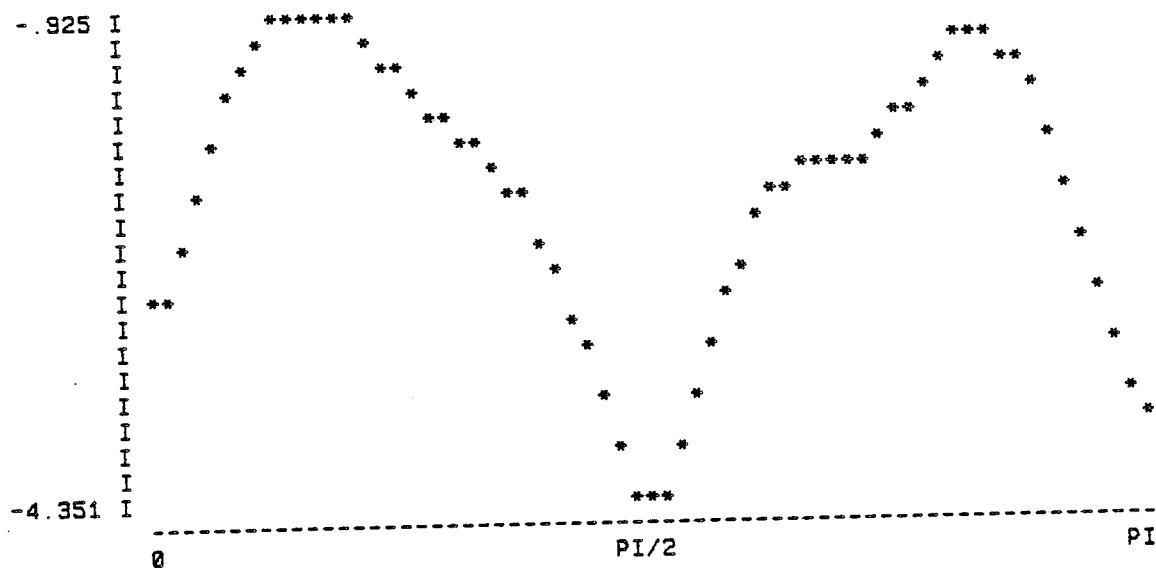


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X59\$LR

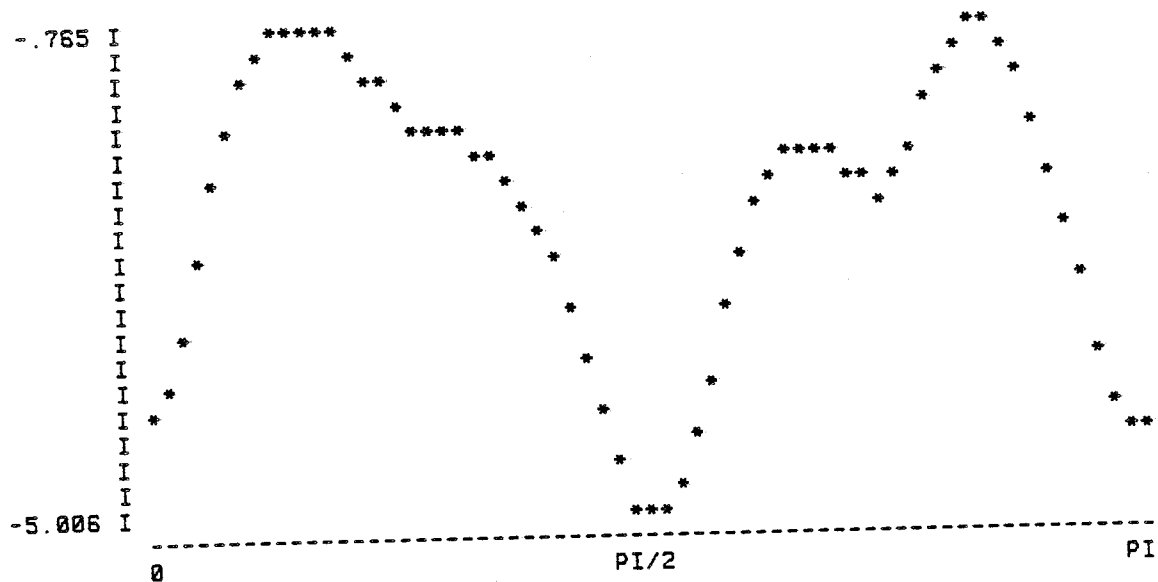


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : X59\$LR

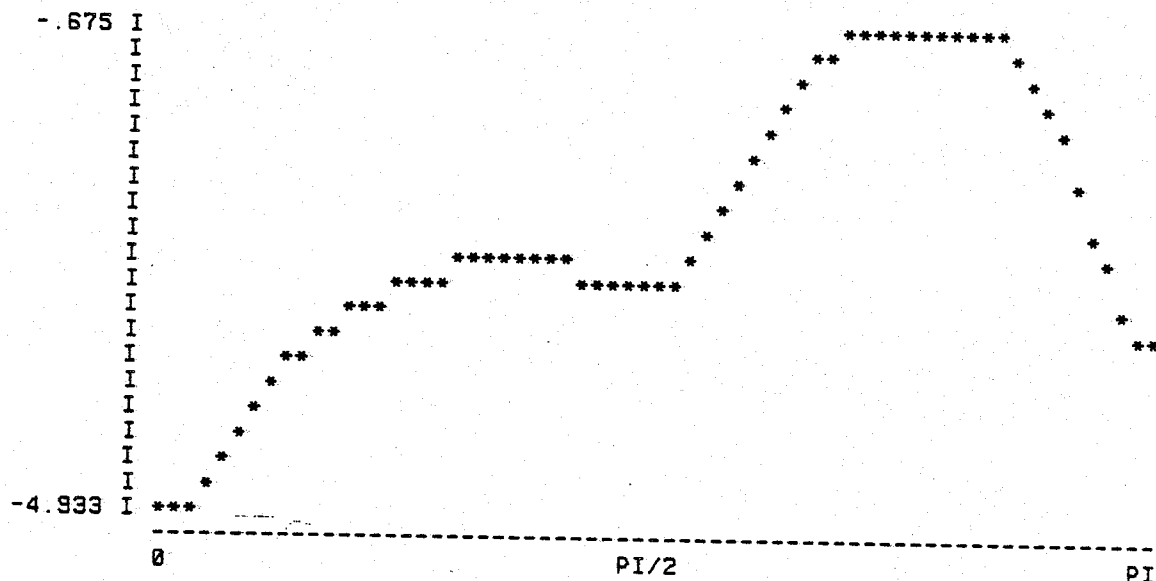


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CNDL

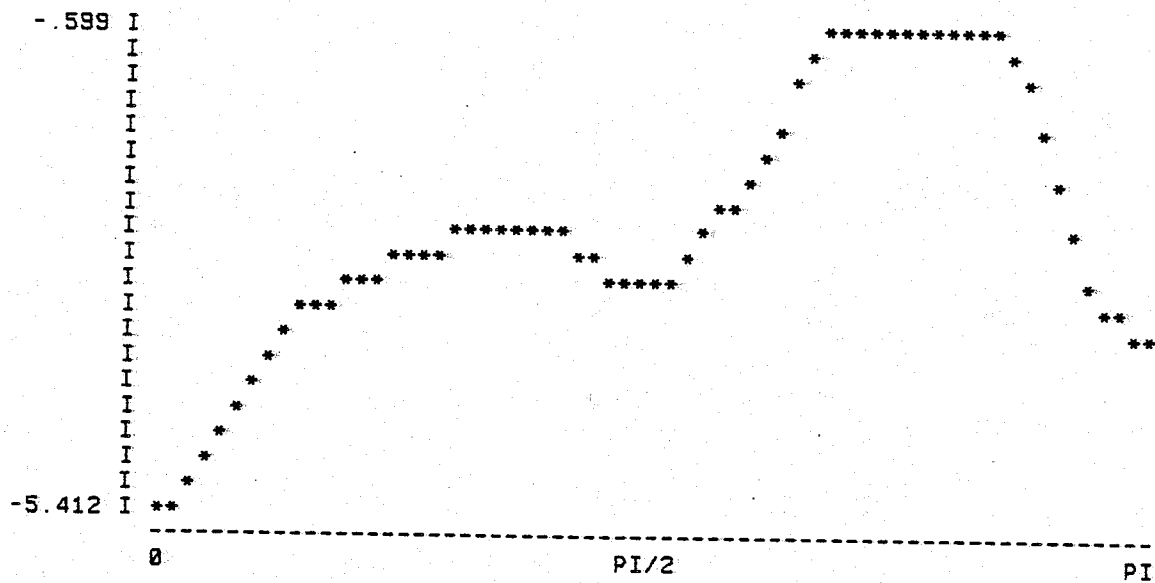


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CNDL

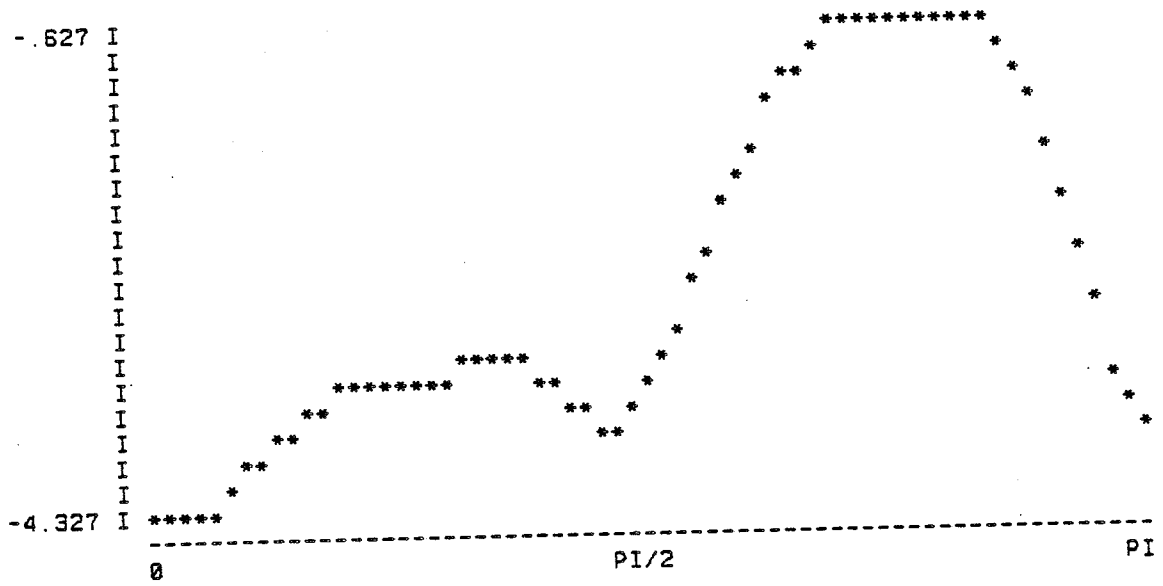


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CNDSLRL

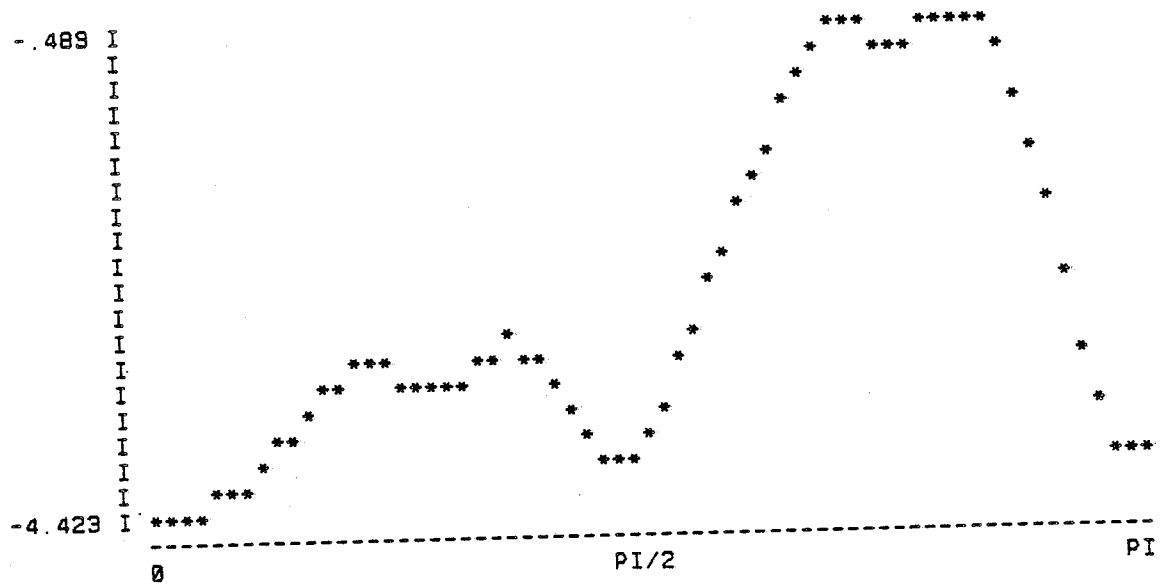


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CNDSLRL



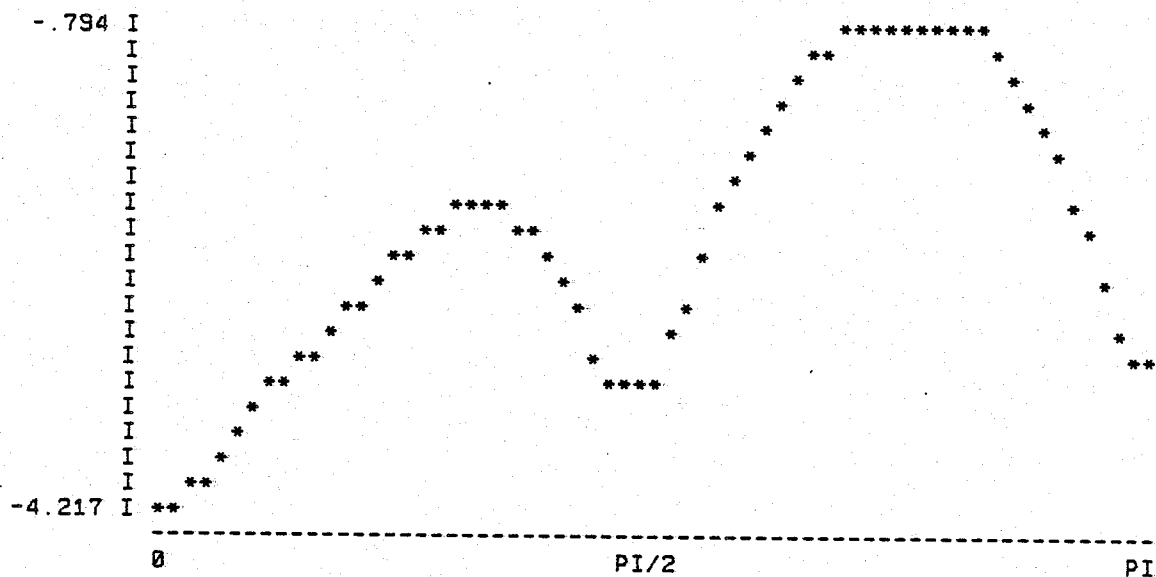
Parzen Window of Length : 30

\*\*\*\*\*



\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CDL



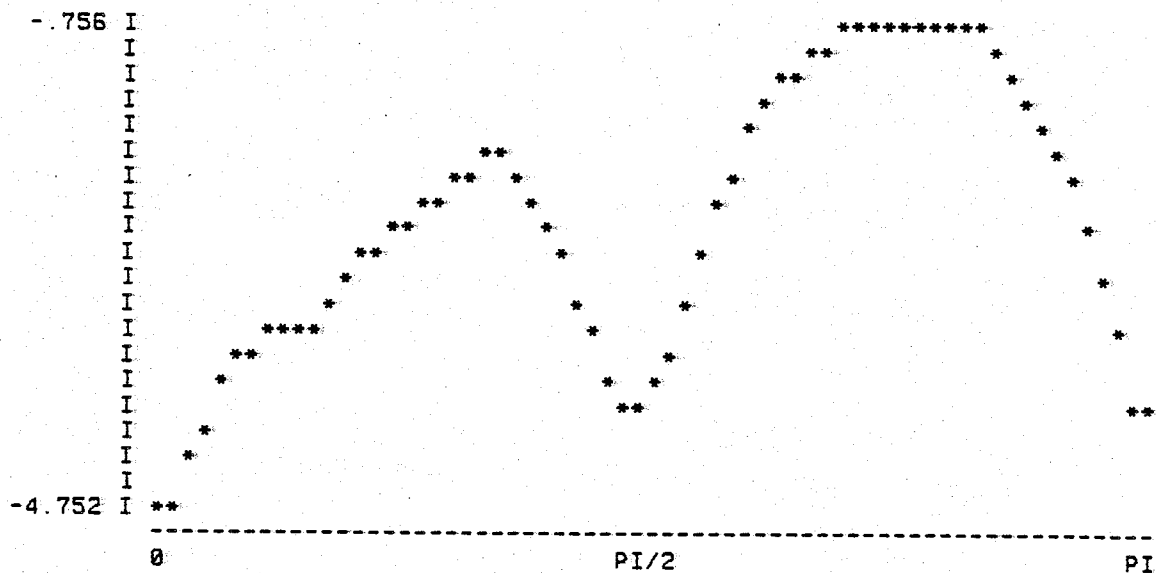
Parzen Window of Length : 20

\*\*\*\*\*

4

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CDL

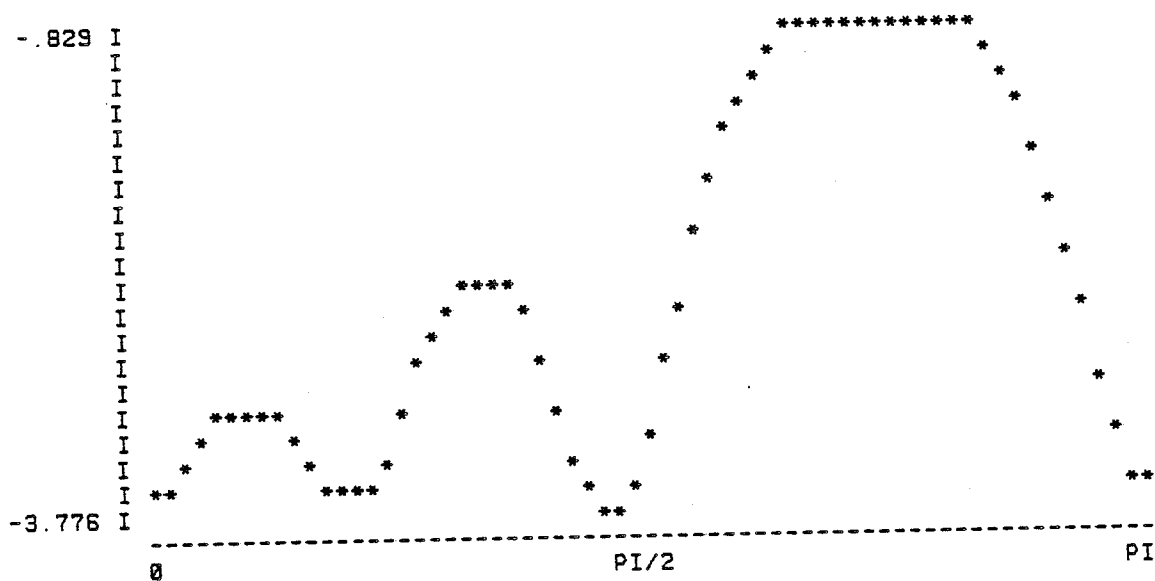


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CDLR10

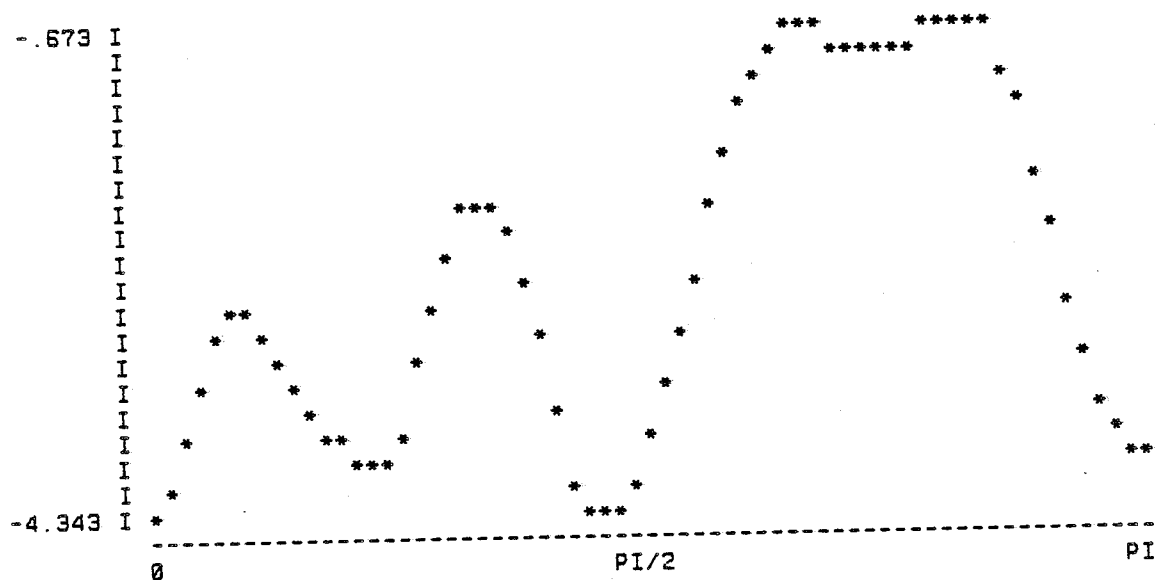


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : CDLR10

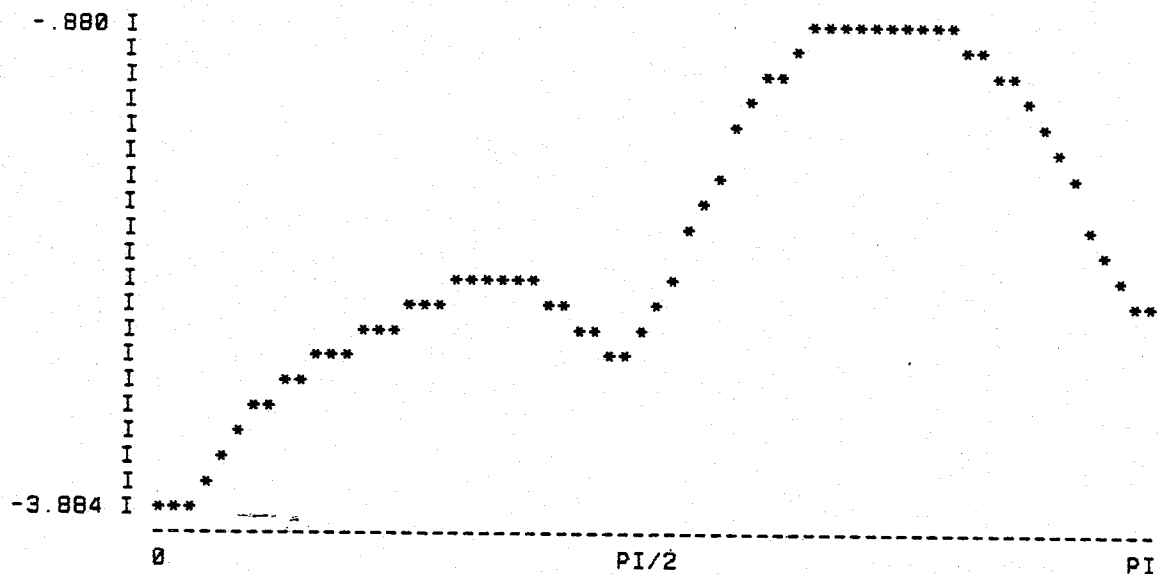


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFEL



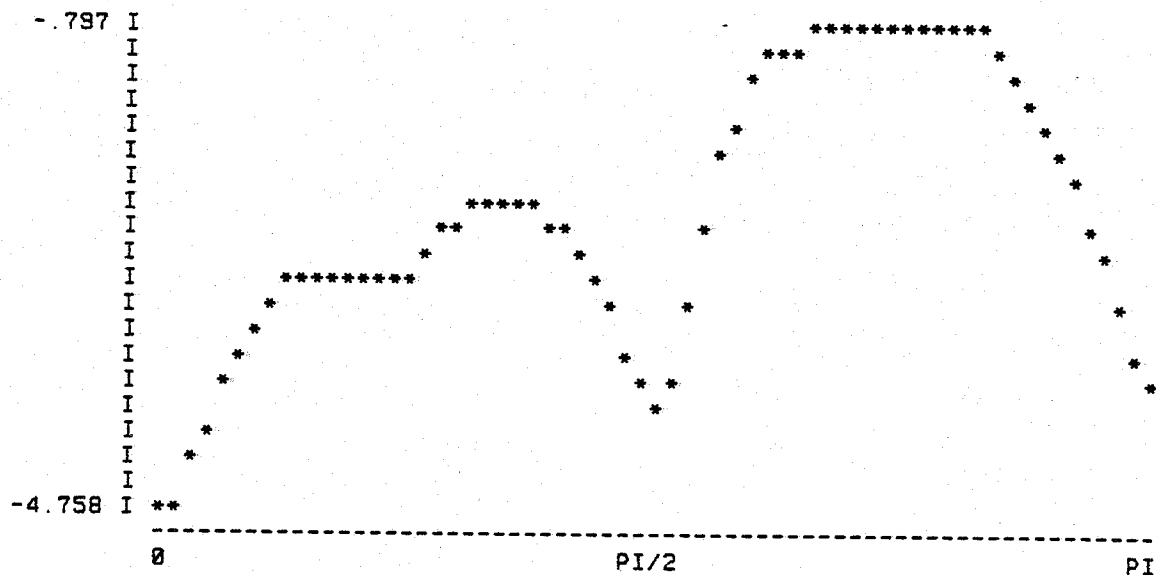
Parzen Window of Length : 15

\*\*\*\*\*

4

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFEL

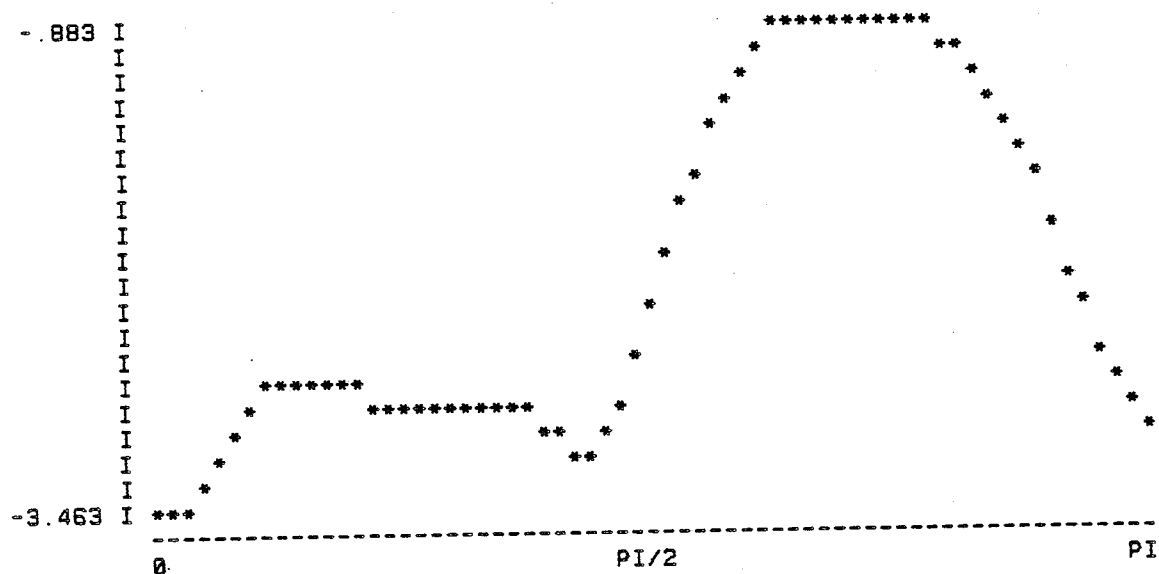


Parzen Window of Length : 25

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFELR

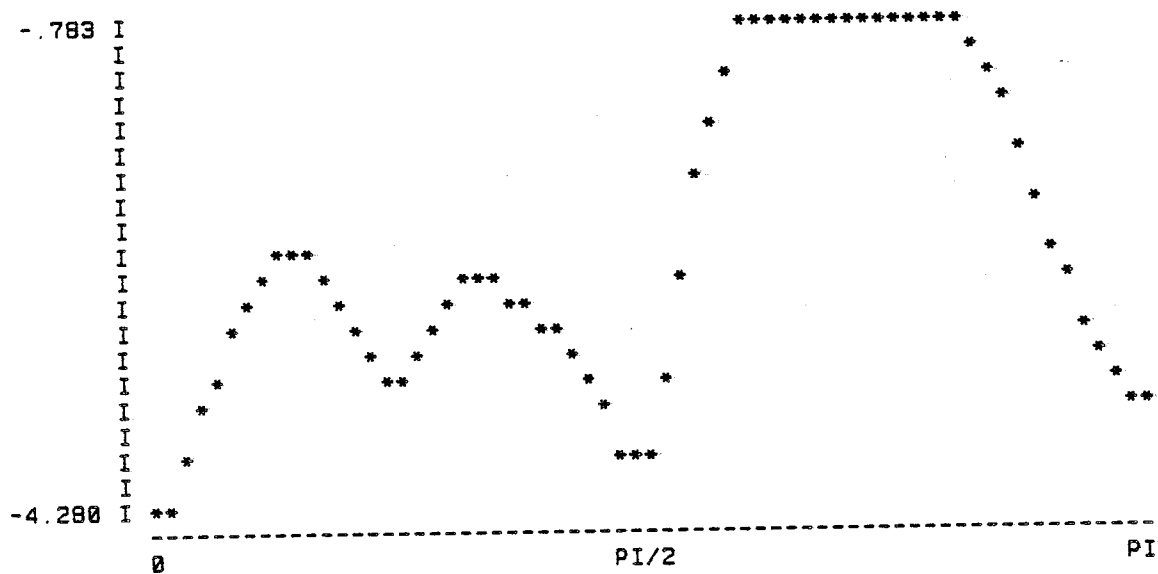


Parzen Window of Length : 15

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFELR

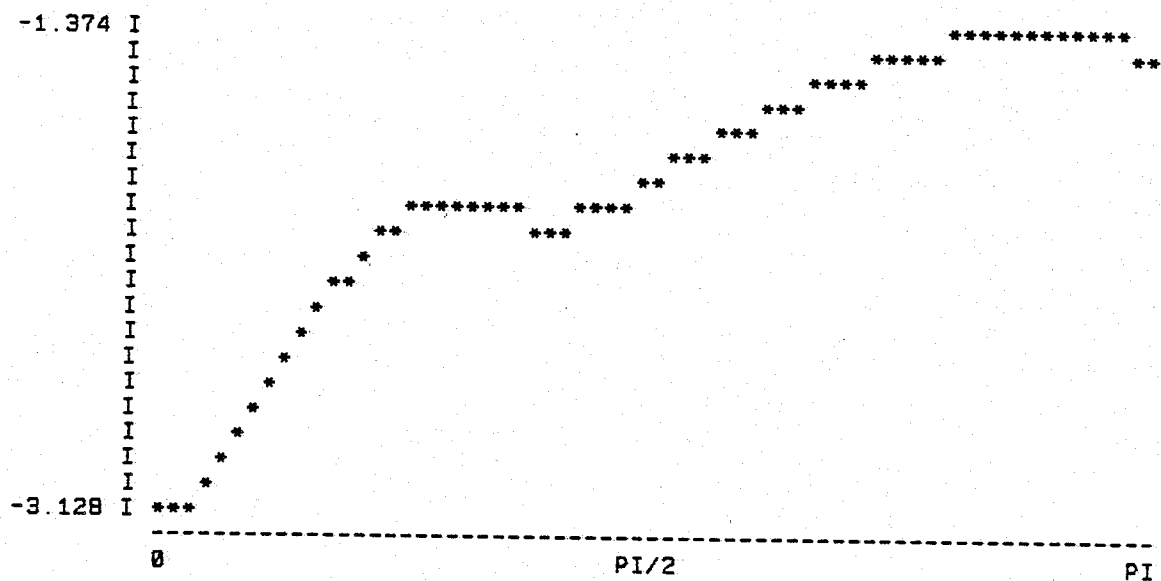


Parzen Window of Length : 25

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFCL

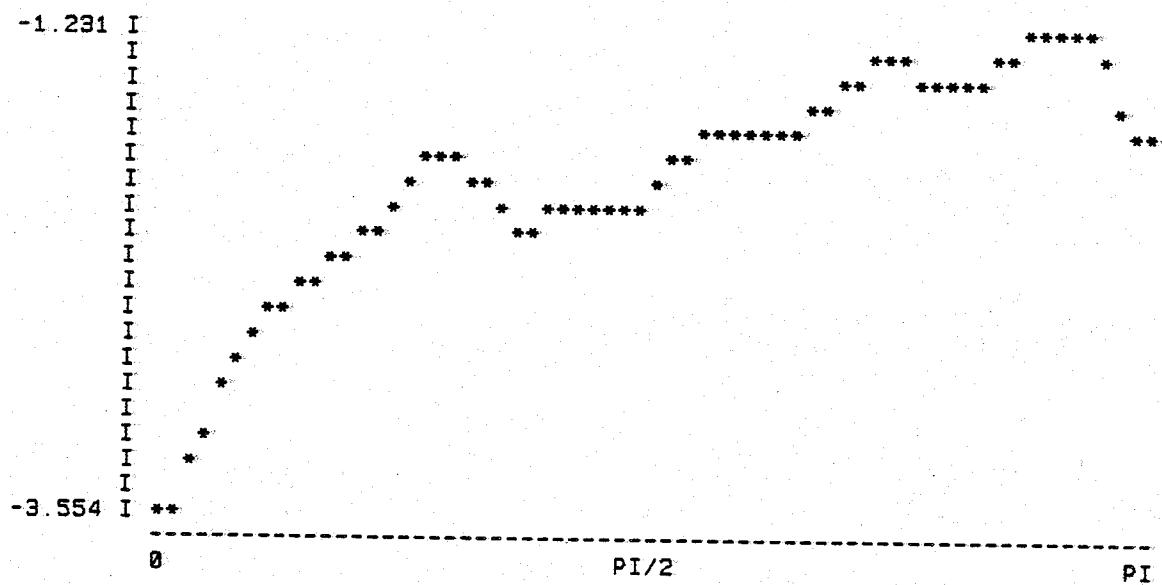


Parzen Window of Length : 15

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : IFCL



Parzen Window of Length : 25

\*\*\*\*\*

The figure is a scatter plot with the following characteristics:

- Y-axis:** Labeled "PI" at the top right. The scale ranges from -2.794 at the bottom to -1.532 at the top, with major tick marks every 0.1 units.
- X-axis:** Labeled "PI/2" at the bottom right. The scale ranges from 0 at the left to 1.0 at the right, with major tick marks every 0.1 units.
- Data Points:** Represented by asterisks (\*). They form a smooth, increasing curve that starts at approximately (0, -2.794) and ends at approximately (1.0, -1.532).
- Grid:** A dashed grid is present, with vertical lines corresponding to the x-axis ticks and horizontal lines corresponding to the y-axis ticks.

\*\*\*\*\*

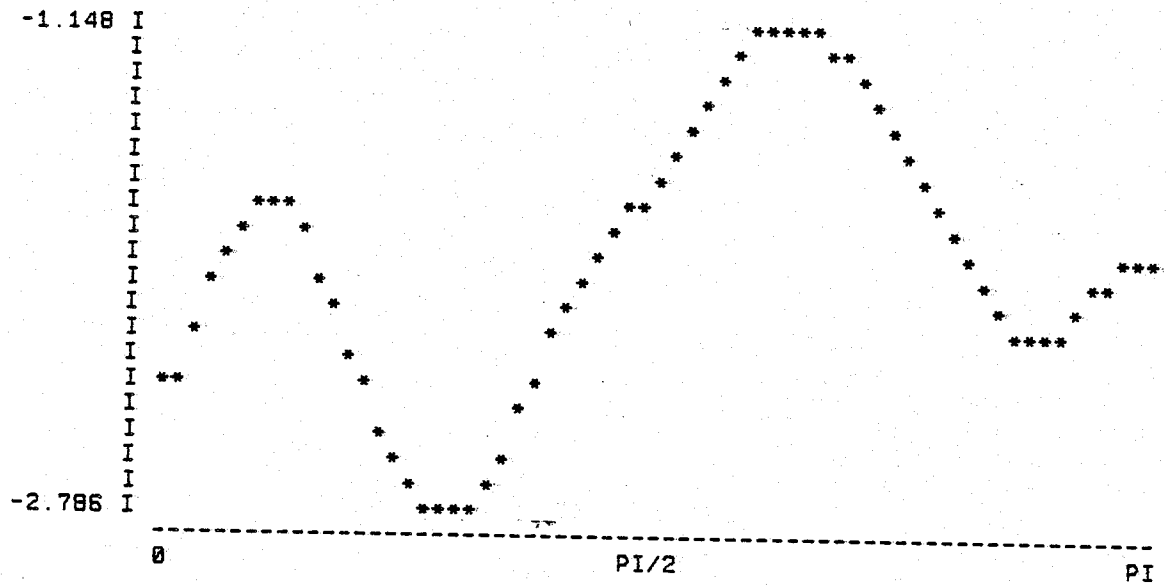
The figure is a scatter plot with the following characteristics:

- Y-axis:** Labeled with values  $-1.407$  at the top and  $-3.013$  at the bottom. The axis is marked with vertical tick marks.
- X-axis:** Labeled with  $0$  at the origin and  $PI$  at the far right. The label  $PI/2$  is positioned below the axis line.
- Data Points:** Represented by asterisks (\*). The points start near  $x=0$  and follow an upward-sloping, slightly irregular path towards  $x=PI$ .
- Reference Line:** A dashed horizontal line is drawn across the plot at the  $y = -3.013$  level.

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : DRSEC

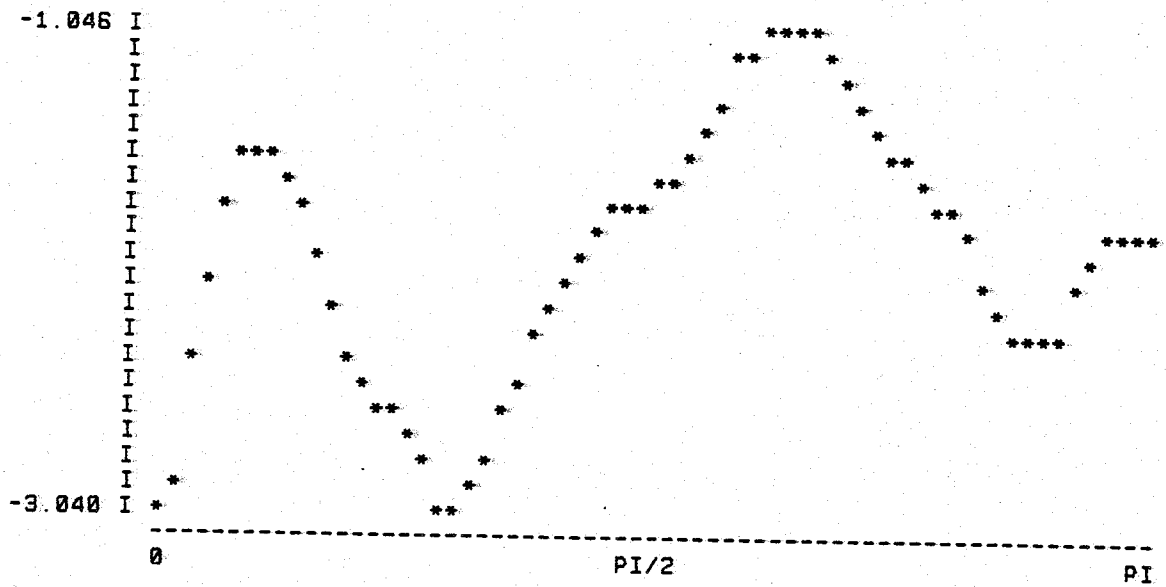


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : DRSEC

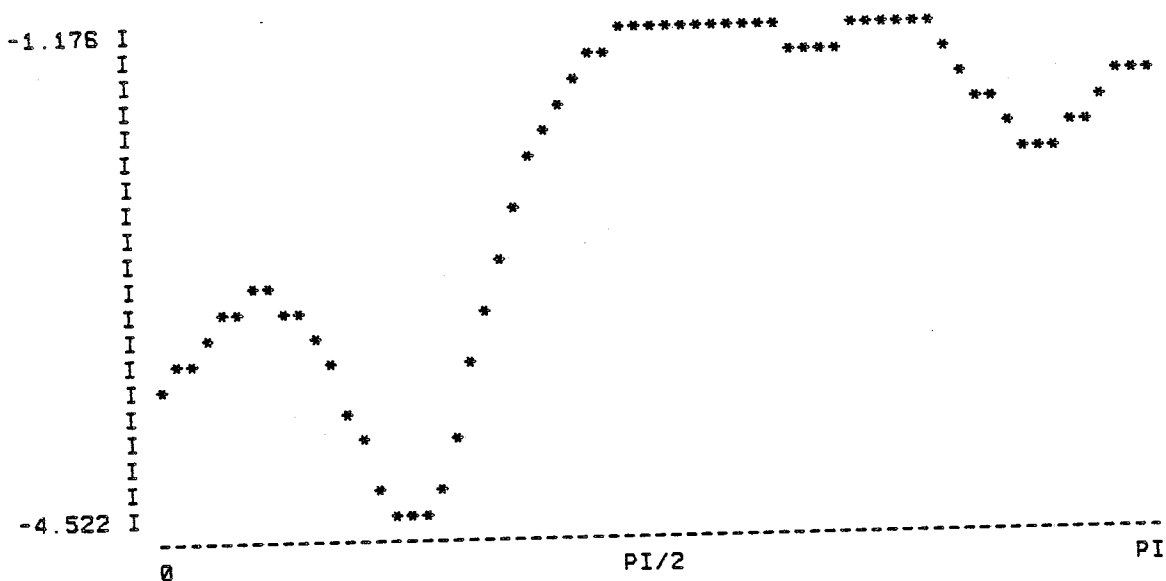


Parzen Window of Length : 30

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : DRSECS

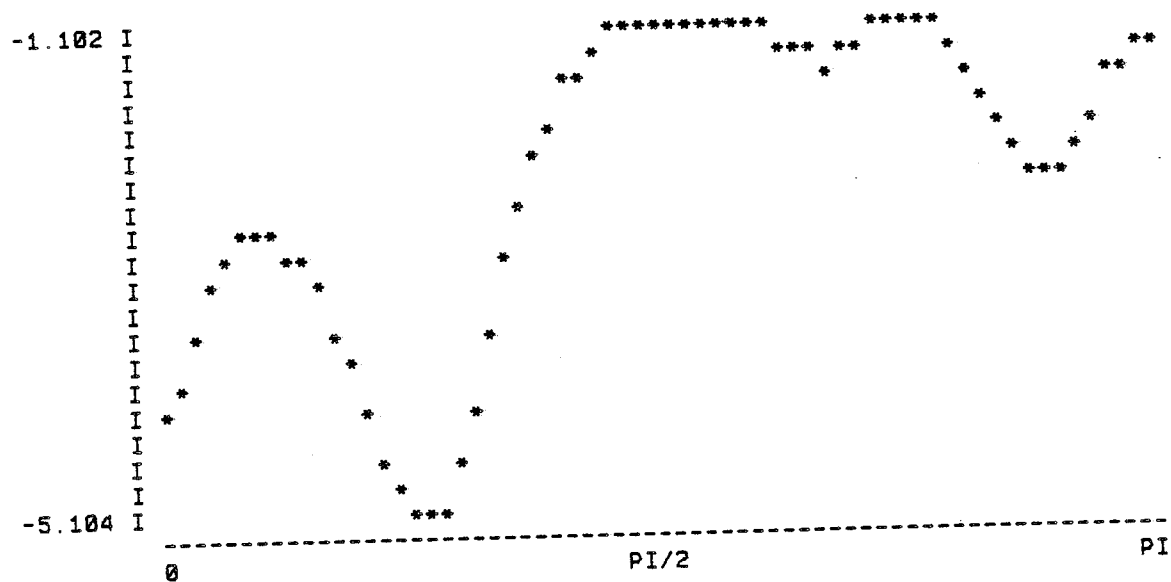


Parzen Window of Length : 20

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : DRSECS



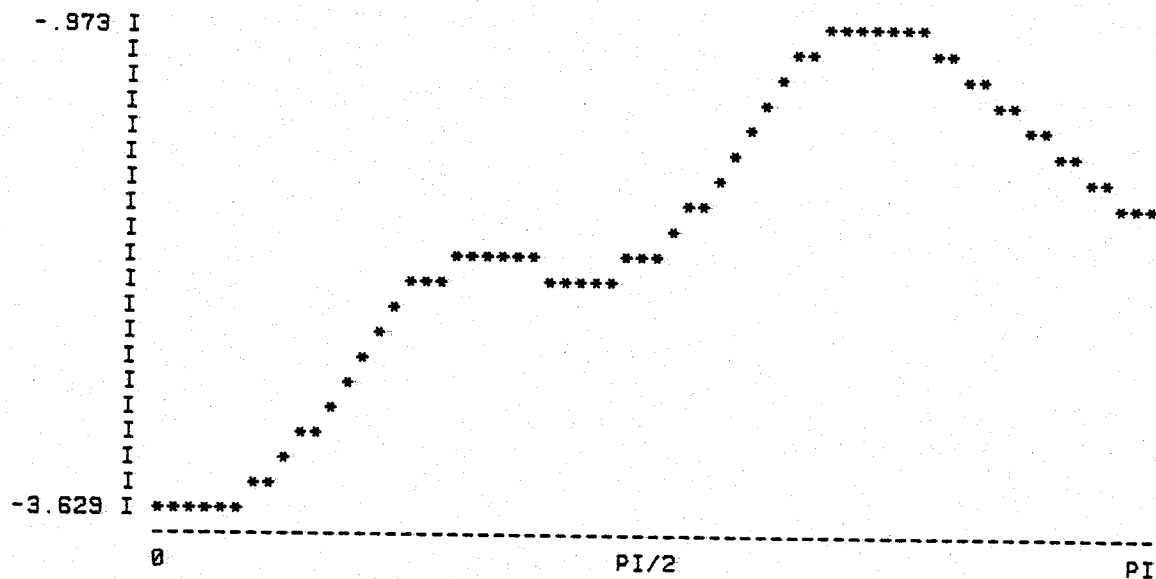
Parzen Window of Length : 30

\*\*\*\*\*



\*\*\*\*\*

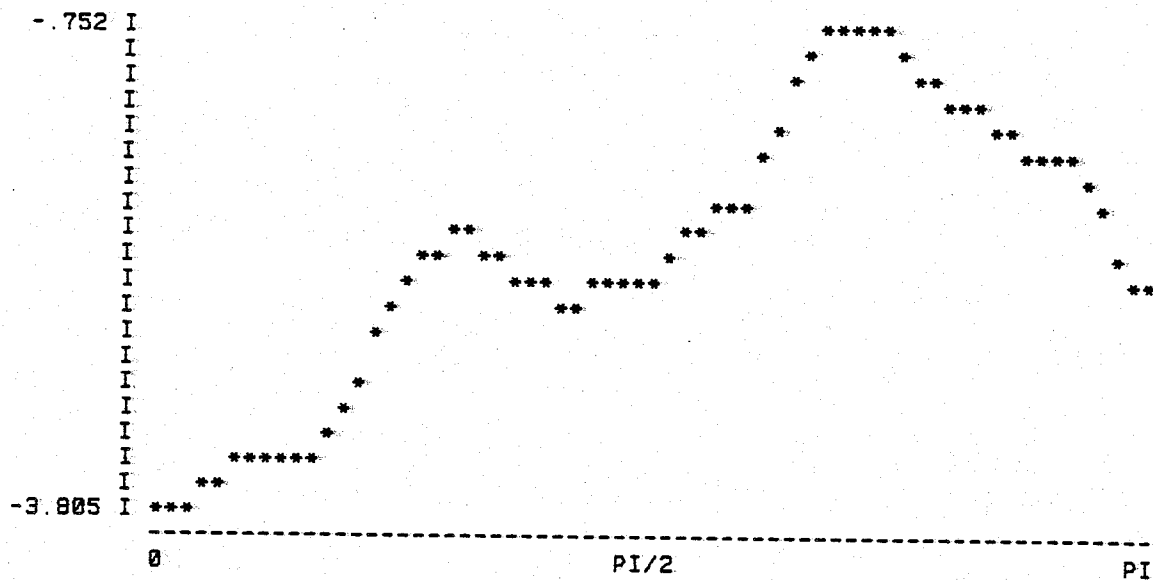
LOGGED NORMALIZED SPECTRUM OF : P6DPL



\*\*\*\*\*

\*\*\*\*\*

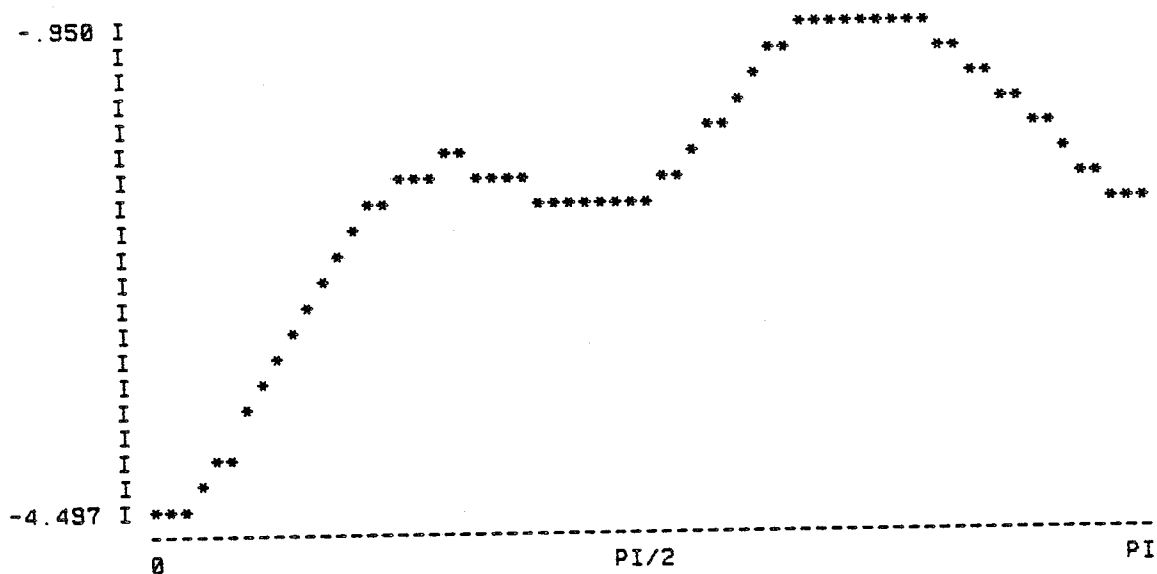
LOGGED NORMALIZED SPECTRUM OF : P6DPL



\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : PGDPL5

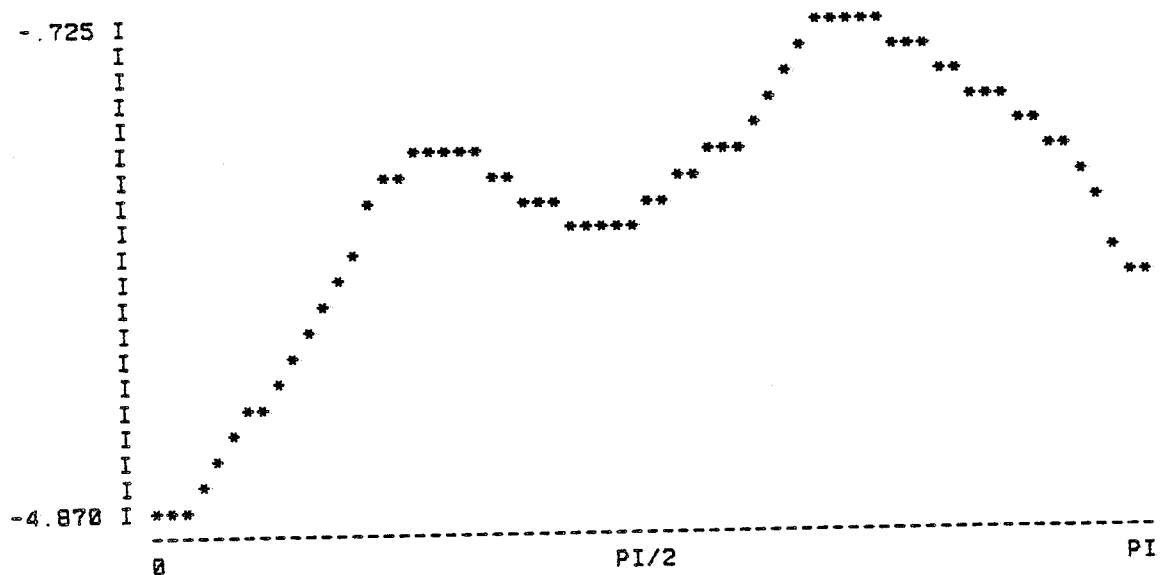


Parzen Window of Length : 15

\*\*\*\*\*

\*\*\*\*\*

LOGGED NORMALIZED SPECTRUM OF : PGDPL5



Parzen Window of Length : 25

\*\*\*\*\*